# IDENTIFICATION AND PREDICTION OF INTER-INDIVIDUAL DIFFERENCES IN COGNITIVE TRAINING TRAJECTORIES: A GROWTH MIXTURE MODELLING APPROACH

Anna P Wolf

University of Tasmania

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy (Clinical Psychology) in the School of Psychology.

June 2015

# ABSTRACT

There is emerging evidence of inter-individual differences in cognitive training responsiveness. Conventional statistics do not adequately address heterogeneity and longitudinal performance trajectories.

Generalised growth mixture modelling (GGMM; Muthén, 2004) was utilised to identify and predict heterogeneous longitudinal cognitive performance trajectories following training. Specific and generalised effects of training were examined. Baseline characteristics such as age, sex and proxies for cognitive reserve were also explored as predictors of trajectories.

Data from 315 community-dwelling older adults (age 55–85 years) from the Active Cognitive Enhancement (ACE) Program training study were analysed. Short-term (VM) and long-term verbal memory (LTVM) and executive functioning (EF) were tested using the Rey Auditory Verbal Learning Test (RAVLT) and the CogState Ltd Groton Maze Learning Test (GMLT) at baseline and at 3-, 6- and 12-month follow-ups.

Generalised growth mixture modelling demonstrated High, Moderate, and Low performance classes for memory performance. High and Low classes were identified for executive function. Also identified were demonstrable performance trajectory gains in the trained individuals of the Low class for executive function, those performing at a low normative level at baseline (Cohen's $d$ = 2.23). These results offer a novel contribution to the literature.

Gains by those trained in the Low performing VM and LTVM classes' performance trajectories were also shown (Cohen's $d$ = 4.48 and 1.38, respectively). However, the

experimental participants were compared to a small number of controls (*n* = 2) thus no meaningful training effects on memory were identified. The GGMM models therefore demonstrated that the multidomain ACE cognitive training program produced some generalised cognitive improvement in healthy older adults, albeit to limited extent.

Age and estimated premorbid IQ (a proxy for cognitive reserve) predicted Low EF performance trajectories compared to High class performances. Trained individuals were more likely to be older and have lower levels of estimated pre-morbid IQ. Individuals who demonstrated executive function performance gains were less likely to demonstrate verbal memory trajectory gains. These findings suggest distinct responses to training in different cognitive domains and/or distinctive inter-individual responses to elements of the multi-domain training program. Caution with interpretation of GGMM labels and predictive factors identified is necessary, given their relativity to the cohort. With this approach, current theories including compensation, magnification, 'Use It or Lose It', plasticity, flexibility, and cognitive reserve are supported. Application of GGMM can also further facilitate development of individually tailored and cost effective cognitive training programs.

## DECLARATION

This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, and to the best of the my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

Signed: Date: 12/1/2016

## STATEMENT OF AUTHORITY OF ACCESS

This thesis may be made available for loan and limited copying in accordance with the *Copyright Act of 1968.*

Signed: Date: 12/1/2016

# ACKNOWLEDGEMENTS

Firstly, I'd like to thank the Australian Research Council (ARC), in collaboration with Alzheimer's Australia, Tasmania (AATas), Department of Health and Human Services (DHHS), and the School of Psychology, University of Tasmania (UTas) for their generous funding support to implement this ambitious research project.

To my primary supervisor, Jeff, thank you for taking me on as one of your students and your encouragement to do a PhD. Thank you for sharing with me some of your immense academic wisdom. I am so grateful for your instrumental assistance in seeing me through the inevitable (multiple) hurdles along the way. Thank you also for your fantastic anecdotes about university life, research and your travels, which always made supervision such a pleasure.

To my co-supervisor Rapson, thank you for your enthusiasm when coming on board the research team, and for introducing me to, and assisting me with implementing growth mixture modelling. Without this assistance, this PhD could not have come about.

To my second co-supervisor Mathew, thank you for your assistance with the study's development, and your clinical expertise and guidance during the running of the program. I'm also very appreciative of your thorough and useful feedback about the final document.

To Sarah, thank you for not only being instrumental in the development and implementation of the ACE project as a whole, but for your professional support and your caring friendship.

To Malcolm, for being such a committed ACE developer, facilitator and instigator of those much needed coffee meetings!

To Avril, thank you for your tireless research assistant work and your generous hospitality in Tasmania.

To Amy, for your fantastic editing assistance. In particular thank you for being so thorough and generous with your time and effort.

To all those who participated in the research study, as well as all the volunteers who devoted their time to the ACE project – thank you all for your integral contributions, enthusiasm and commitment.

To my treasured friends and family both near and far. I love you all dearly. You kept me going! I can't wait to have loads more spare time now to continue the fun!

To Anastasia, thank you for your friendship, thoughtfulness and crazy ways. You do make me laugh!

To Leonie, thanks for all the fun times, "walkie talkies", and of course for our continuing friendship for all these years.

To Dani - Wifey! Thank you for sharing the "nanna rug" time, and for being such a honest and caring friend.

To Sinead, thank you for our fantastic long chats at Uni, Alzheimer's, cafes and via Skype! I look forward to more!

To Vandita, for your friendship in both Tassie and Sydney. Just hearing your voice instantaneously fills my heart with joy!

To Eddy, for welcoming me so openly to Hobart, for your assistance with surviving the “smelly boys” and for continuing to share your life with me.

To Dave, Buddy, thank you for always being such a loving brother! I miss you now you’re in Canada, but I’m so happy you’re so happy over there. I know you’ll always be there for me.

To DiDi, thank you for your open ear and being such a nurturing aunty.

To Alan, thanks for our stimulating chats about movies, politics and for sharing your amazing cooking and incredible travel photos!

Joan for being such a kind, caring and immensely compassionate person. I am so grateful for all your unyielding support and understanding. I’m also so thankful for you sharing your home with me and for our family dinners!

Finally to my extraordinary parents, Penny and John. Words cannot fully encapsulate how much gratitude I have for your undying love, support, dear friendship and mentorship all the way through this long journey. Mum, Dad, what can I say? I simply couldn’t have done it without you.

# ABBREVIATIONS

| | |
|---|---|
| 3MS | Random-effects regression on performance on the modified MMSE |
| AACD | Age-associated cognitive decline |
| AAMI | Age-associated memory impairment |
| AATas | Alzheimer's Australia, Tasmania |
| ABIC | Sample-size Adjusted Bayesian Information Criterion |
| ABS | Australian Bureau of Statistics |
| ACE | Active Cognitive Enhancement |
| ACTIVE | Advanced Cognitive Training for Independent and Vital Elderly |
| AD | Alzheimer's disease |
| Adjusted LRT | Lo-Mendell-Rubin Adjusted likelihood ratio test |
| ANOVA | Analysis of variance |
| APA | American Psychiatric Association |
| ARCD | Age-related cognitive decline |
| BHQ | Brain health questionnaire |
| BIC | Bayesian information criterion |
| BRC | Brain reserve capacity |
| CASP-12 | Control, Autonomy, Self-realisation and Pleasure scale |
| CBT | Cognitive behaviour therapy |
| CC75C | Cambridge City Over 75 Cohort study |
| CIND | Cognitive impairment, no dementia |
| CPAL | Continuous paired associate learning task |
| CR | Cognitive reserve |
| DRS-2 | Dementia Rating Scale (2nd edition; DRS-2) |
| DASS-21 | 21-item version of the Depression, Anxiety and Stress Scale |

| | |
|---|---|
| DET | Detection task |
| DSM-IV-TR | Diagnostic and Statistical Manual, 4th Edition, Text Revision |
| EE | Environmental enrichment |
| EEG | Electroencephalography |
| EF | Executive function |
| FIML | Full information maximum likelihood |
| fMRI | Functional magnetic resonance imaging |
| FSIQ | Full Scale Intelligence Quotient |
| GBGM | Group-based growth modelling |
| GBM | Group-based modelling |
| GGMM | Generalised growth mixture modelling |
| GMLT | Groton Maze Learning Task |
| HADS | Hospital Anxiety and Depression Scale |
| HAROLD | Hemispheric Asymmetry Reduction in Older Adults |
| Health ABC | "Health, Ageing and Body Composition" study |
| HVLT | Hopkins Verbal Learning Test |
| HVLT-R | Hopkins Verbal Learning Test, Revised |
| ICV | Intra-cranial volume |
| IDN | Identification task |
| LBLS | Long Beach Longitudinal Study |
| LCGA | Latent class growth analysis |
| LEQ | Lifetime of experiences questionnaire |
| LGCM | Latent growth curve modelling |
| LGM | Latent growth modelling |
| LMR | Vuong-Lo-Mendell-Rubin likelihood ratio test |

| | |
|---|---|
| LTVM | Long-term verbal memory |
| MAR | Missing at random (data) |
| MCI | Mild cognitive impairment |
| mMMSE | Modified Mini-Mental State Examination |
| MMQ | Multifactorial memory questionnaire |
| MMSE | Mini Mental State Examination |
| MRC CFAS | Medical Research Council Cognitive Function and Ageing Study |
| NHMRC | National Health & Medical Research Council |
| OCL | One card learning task |
| Octo-Twin | Origins of Variance in the Oldest-Old: Octogenarian Twins Study |
| ONB | One back task |
| PET | Positron emission tomography |
| RAVLT | Rey Auditory Verbal Learning Test |
| *SD* | Standard deviation |
| *SE* | Standard error |
| SEM | Structural equation modelling |
| SLS | Seattle Longitudinal Study |
| TWOB | Two back task |
| VM | Verbal memory |
| VLS | Victoria Longitudinal Study |
| WHO | World Health Organization |
| WMS | Wechsler Memory Scale |
| WTAR | Wechsler Test of Adult Reading |

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# Chapter 1

# Introduction

Currently there is limited research investigating inter-individual differences in longitudinal cognitive performance trajectories following cognitive training in older adults. This thesis and the associated empirical study extend beyond past research by identifying inter-individual differences in training outcomes and exploring their predictors through the use of a relatively new statistical technique – generalised growth mixture modelling (GGMM; Muthén, Brown, Khoo, Yang, & Jo, 1998; Muthén & Muthén, 1998–2010; Muthén & Shedden, 1999). Specifically, there was identification of inter-individual responsiveness to training in cognitive domains shown to decline with age, as well as exploration of individual baseline characteristics predicting these heterogeneous cognitive performances. The thesis subsequently explores neuropsychological and neurobiological theories such as the 'Use it or lose it' hypothesis, plasticity, flexibility, and cognitive reserve (CR) that dominate the ageing literature. The current study was therefore both theoretically and statistically motivated.

## Cognitive Ageing: Age-related Cognitive Decline and Successful Ageing

Ageing is considered to be the process by which we get older, which, of course, begins at birth. Scientifically however, it is considered to be the process by which we decline following maturity. This is also referred to as senescence (Blackburn, 2007). From a cognitive perspective, ageing is generally associated with robust, progressive deficits in cognitive performance. Cross-sectional and longitudinal comparisons consistently reveal that increases in age are associated with a decline in performance on a wide range of cognitive measures, including verbal memory (VM), processing speed, visuo-spatial skills, attention and executive functions, including working memory (Buckner, 2004; Deary, 2009; Hedden & Gabrieli, 2004; Schaie, 2000). Whilst there is considerable variation in the literature around the terms used to define cognitive function that is less than optimal (but non-pathological) in older adults, (as will be discussed later in the thesis), normal cognitive declines can be, and often are, referred to as *Age-related Cognitive Decline* (ARCD).  This followsthe Diagnostic and Statistical Manual, 4$^{th}$ Edition, Text Revision (DSM-IV-TR; American Psychiatric Association [APA], 2000) definition.

One of the most striking aspects of many investigations into ARCD are the large inter-individual differences in the extent that people are affected by ARCD (Barnes et al., 2007; Lindenberger & von Oertzen, 2006; Nelson & Dannefer, 1992; Schaie, 1994). The literature refers to individuals who are 'successfully ageing'. This, in a cognitive context, incorporates a minority of individuals who do not exhibit large decrements in cognitive functioning in domains typically seen in epidemiological, cross-sectional and longitudinal studies (Barnes et al., 2007; Carey, 2007; Jones et al., 2005; Lindenberger & von Oertzen, 2006; Nelson & Dannefer, 1992; Raz, 2009; Schaie, 1994; Yaffe et al., 2009).

**Ageing Population**

The world is currently experiencing global population ageing at an unprecedented rate and thus, the impact of ARCD will also increase. The total number of people aged 65 and older worldwide is predicted to dramatically increase, from 690 million in 2010 to nearly 1.5 billion by 2050. In addition to increases in absolute numbers, it is expected the proportion of people ≥ 65 years of age will nearly double from 9% to 16% (United Nations, 2009). Australia is also currently witnessing an increase in the ageing population (Australian Bureau of Statistics [ABS], 2009). In 2007, people aged ≥ 65 years made up 13% of Australia's population and this proportion is projected to increase to between 23% and 25% in 2056 (ABS, 2008). Driving this trend are two major factors: increased life expectancy at birth, including sustained low levels of fertility, and a large cohort of baby boomers (ABS, 2008, 2009; Access Economics, 2001). The massive baby boom cohort (the 5.5 million individuals born post-World War II, *i.e.*, between 1946 and 1965) is also getting older, placing Australia at a critical turning point with acceleration of the impacts of ageing (Quine & Carter, 2006).

**Consequences of an Ageing Population**

Population ageing has many concerning economic and psychosocial consequences. Although ARCD does not profoundly affect real-world function, the cognitive domains affected are nonetheless important for carrying out everyday activities and other cognitively demanding tasks, such as living independently and leading an engaging and fulfilling life. An ageing population also creates greater demands on health care, an increased need for residential support (*e.g.*, in-home assistance, residential care or a move into a family member’s home) and an added caregiver burden (Fillenbaum, 1985; Fogel, Hyman, Rock, & Wolk-Klein, 2000). There will be a mass efflux of older adults out of the workforce, therefore increasing the number of people accessing government guaranteed and/or funded superannuation and pension plans. Access Economics (2001) estimates that the economic cost of ageing will be approximately $16–24 billion (in 2000 dollars) by 2031. When considering further down the trajectory of declining cognitive performance, it is estimated that delaying the cognitive onset of Alzheimer's disease (AD) by five years, would create a 49% reduction in the total number of cases projected by mid-century (Access Economics, 2004).

At a psychosocial level, overt signs of cognitive decline are amongst the most distressing aspects of ageing. Memory in particular is one of the first abilities to show decline and is widely considered the characteristic trait of ageing in our culture. It is one of the most common complaints of older people (Carey, 2007; Kramer & Willis, 2002; Thompson & Foth, 2005). In fact, nearly half of community-dwelling individuals aged ≥ 60 years express concern about declining mental abilities (Ball, Edwards, Ross, McGwin, 2010). The relationship between increasing memory deficits and dementia is well established (Thompson & Foth, 2005). Age-related cognitive decline also contributes to a loss of autonomy and the ability to engage in social and in recreational pursuits, and can produce negative consequences on quality of life (Greene & Williams, 1996; Mahncke, Bronstone, & Merzenchi, 2006a). Individuals with cognitive decline are also at increased risk of progressing to mild cognitive impairment (MCI) and dementia (*e.g.*, AD), further affecting their functional capacity, autonomy and quality of life (Amieva et al., 2005; Plassman, Williams, Burke, Holsinger, & Benjamin, 2010).

With these numerous economic and psychosocial consequences, there is a significant need to explore why some individuals demonstrate less ARCD. This may then offer insights into possible avenues in which these factors can be utilised or manipulated. Intervention may allow for maintenance of current levels of cognitive performance, reduce the rate of ARCD or, ideally, increase the number of individuals exhibiting *successful ageing* – those who appear to be withstanding the effects of ageing on cognition (Carey, 2007; Jones et al., 2005; Lindenberger & von Oertzen, 2006; Raz, 2009; Schaie, 1994; Yaffe et al., 2009).

## Theories Explaining Inter-individual Differences in Age-related Cognitive Decline

There are a numerous similar, linked and overlapping theories and constructs across the scientific literature with regard to the influences on the manifestation of inter-individual differences in the expression of ARCD and in successful ageing. Of importance in the neuropsychological literature are the concepts of cognitive reserve (CR)[1] and the related neurobiological concept of plasticity, in addition to what can be considered a central, overarching and predominating theory, popularly known as 'Use it or lose it'. With their numerous specific nuances, this cluster of concepts and theories ultimately highlight the potentially cognitively-enriching effects of environmental experience (Tucker-Drob & Salthouse, 2011). This influence of environmental experience is applied here through specific, active, intentional manipulation of older adults' environments via cognitive training intervention to maintain or improve cognitive performance.

## Cognitive Training

The cognitive training literature demonstrates mixed results as to the extent and nature of effectiveness of cognitive training interventions on cognitive performance in older adults. Evidence from some studies suggest that cognitive training improves cognitive functioning in older people, with some group data showing promising results, although effect sizes vary (Martin, Clare, Altgassen, Cameron & Zehnder, 2011; Papp,

[1] The operatizational definition will be further detailed in Chapter 2

Walsh. & Snyder, 2009; Valenzuela & Sachdev, 2009). There are also limited studies demonstrating generalised effects of training – effects that represent maintenance or improvement in cognitive domains beyond those specifically trained. More generalised effects of cognitive training must be considered when assessing the efficacy of cognitive training more broadly and the legitimacy of the 'Use it or lose it' and associated theories. Similarly, few studies conduct adequate long-term follow-up, nor trajectories of performance across time. Given that the ultimate goal of cognitive training should be to instigate lasting cognitive change, it is imperative to look at longer-term effects of training outcomes.

**Individual Differences in Cognitive Training Responsiveness**

Importantly, there is emerging evidence of inter-individual differences in cognitive training responsiveness. In fact, inter-individual differences can be quite large (Bissig & Lustig, 2007; Yesavage, Sheikh, Tanke, & Hill, 1988). Ignoring this heterogeneity obscures the accuracy of training study results and raises questions about the validity of past research conclusions that do not account for these differences (Boron, Turiano, Willis, & Schaie., 2007b; Duncan et al., 2002; Fairchild, Friedman, Rosen, & Yesavage, 2013; Fandakova, Shing, & Lindenberger, 2012; Langbaum, Rebok, Bandeen-Roche, & Carlson, 2009; Zelinski et al., 2007).

The existence of inter-individual differences highlights issues with the use of conventional group-level statistics. In addition, conventional statistics are not ideal in the training context because the process of change itself needs to be examined. Most current statistical approaches focus on group effects and 'responders' *vs.* 'non-responders' at a specific time-point. Observing the individual response trajectories provides more information about the rate of change in performance (Baltes &

Nesselroade, 1979; Baltes & Kliegl, 1992; Collins & Horn, 1991; Kliegl, Smith & Baltes, 1989; Langbaum et al., 2009; Muthén, 2004). Considering performance trajectories can show whether a treatment intervention, such as a cognitive training program, can alter an individual's normative growth trajectory that would exist without treatment (Baltes & Kliegl, 1992; Deary et al., 2009a; Hedden & Gabrieli, 2005; Kliegl et al., 1989; Langbaum et al., 2009; Muthén, 2004; Terrera et al., 2010; Nagin & Odgers, 2010). There is, however, a paucity of research considering individual differences and trajectories of cognitive training responsiveness, and mixed levels of evidence of cognitive training benefits in older adults.

**Predictors of Heterogeneity**

Given the heterogeneity of performance in cognitive training outcomes, it is important to identify the predictive characteristics of individuals who do and do not benefit (Nagin & Odgers, 2010; Tucker-Drob & Salthouse, 2011). Individual baseline characteristics – such as age, sex and proxies for CR, including indices of intelligence and education – that represent inherent cognitive capacity and lifetime experiences that may influence an individual's receptiveness to intervention, may predict heterogeneous performance outcomes. Investigating these baseline characteristics may provide further understanding of the effect of such characteristics on training outcomes, leading to a more comprehensive exploration of the 'Use it or lose it' and associated theories by examining how we can predict to whom these theories best apply.

Identifying specific predictors of performance heterogeneity is also of practical use to ensure there is the greatest possible effect for a wider range of individuals and to thereby increase cost-effectiveness of training (*e.g.*, Baldi, Plude, & Schwartz, 1996;

Hastings & West, 2009; Kliegl, Smith & Baltes, 1990; Rebok, Carlson, & Langbaum, 2007; Verhaeghen, Marcoen, & Goossens, 1992; West & Hastings, 2011). Furthermore, exploring specific predictors of inter-individual differences in cognitive training responsiveness may also encourage the development of alternative intervention applications or training designs (*e.g.*, pre-training or booster sessions) for those who show less benefit from existing training programs (Raz, 2009). This is of particular importance when the personal and financial costs of ageing are set to grow substantially (Access Economics, 2001; Papp et al., 2009).

Whilst there have been some studies into the effect of age, sex and proxies of CR on training responsiveness, evidence is mixed. Age is particularly well explored when testing the upper limits of cognitive performance (Hertzog et al., 2009). Proxies for CR, education and indices of intelligence have been discussed and debated; although dominant research and theories suggest they are supportive of training gains, both age and proxies for CR are not always associated with training effects (*e.g.*, Dunlosky & Garrett, Macdonald, & Craik, 2012; Dunlosky & Hertzog, 1998b; Langbaum et al., 2009; Lövdén, Bäckman, Lindenberger, Schaefer, & Schmiedek, 2010; Stern et al., 1994; Sullivan, 1964; West & Tomer, 1989). Furthermore, there are few explorations of the impact of sex on cognitive training and once again effects vary across studies (Herlitz, Nilsson & Bäckman, 1997).

Contributing to the lack of clarity in the literature may be the insufficient number of quality studies. This is particularly the case for studies exploring predictors of inter-individual differences and longitudinal performance trajectories (Herlitz et al., 1997; Martin et al., 2011). While studies addressing these issues are emerging (*e.g.*, Gross et al., 2012; Langbaum et al., 2009), such studies adequately doing so are rare. Furthermore, most studies focus on the cognitive domain of memory.

Thus, the current understanding of the influence of these baseline characteristics on the impact of training outcomes is unclear. Further exploration of distinct cognitive trajectories across time is warranted, particularly through the use of novel and sophisticated statistical techniques.

### Group-based Growth Modelling: Generalised Growth Mixture Modelling

Group-based growth modelling (GBGM) techniques can provide a statistical solution to the identification and prediction of heterogeneity of longitudinal performance trajectories. Group-based growth modelling assumes individuals are members of a finite number of latent classes, or subpopulations (Muthén, 2004). These classes consist of individuals with similar trajectories of performance change over time. Inter-individual differences are shown because individuals across classes are heterogeneous (Muthén & Muthén, 2000; Muthén et al., 2002). One of the most useful GBGM methods is generalised growth mixture modelling (GGMM; Muthén et al., 1998; Muthén & Muthén, 1998–2010; Muthén & Shedden, 1999). Generalised growth mixture modelling is a relatively novel statistical approach that can incorporate covariates into the model. This integrated method allows simultaneous examination of predictor impact on longitudinal trajectories, rather than considering them as outcomes in post hoc comparisons (Connell & Frye, 2006; Muthén et al., 2002, Muthén, 2004; Nagin, 1999; Roeder, Lynch, & Nagin, 1999). The incorporation of covariates into the model is one of the most pertinent aspects of GGMM (Muthén et al., 2002, Muthén, 2004).

Generalised growth mixture modelling is particularly appropriate in the context of the current study, given emerging evidence suggesting inter-individual training responsiveness to cognitive training (Duncan et al., 2002; Hedden & Gabrieli, 2005;

Lindenberger & von Oertzen, 2006; Lövdén et al., 2010; Park, Gutchess, Meade, & Stine-Morrow, 2007; Raz, 2009; Terrera et al., 2010; Yaffe, 2009). It provides a more valid indication of individual performance differences than more conventional statistical methods, which may obscure information regarding the heterogeneity of individual performances within a sample exhibiting robust change (Connell & Frye, 2006; Langbaum et al., 2009; Willis & Schaie, 1987). Generalised growth mixture modelling is also appropriate because it considers the different classes as representing meaningful strata (Kreuter & Muthén, 2007; Nagin & Odgers, 2010) or at least sufficiently so, such that the strata can have real world application. Thus these strata can be used to explore the 'Use it or lose it' and associated theories, and guide clinical decision making when allocating individuals to training programs (Gueorguieva et al., 2007; Kreuter & Muthén, 2007). It was therefore selected as the optimal model from which to draw conclusions in this study.

## Rationale and Aims of the Current Empirical Study

The current empirical study aimed to extend past research. It addressed the lack of clarity with current lines of investigation into the inter-individual effectiveness of cognitive training interventions when conventional statistics are implemented. A relatively new statistical technique, GGMM, was used to explore and predict heterogeneity in longitudinal cognitive trajectories following training. It aimed to utilise GGMM to provide a clearer exploration of inter-individual responsiveness to training and prominent constructs such as the 'Use it or lose it' and associated theories.

The study had three central aims:

1. To investigate inter-individual differences in training effects by comparing longitudinal cognitive performance trajectories of individuals who undertook

training with the performance trajectories of those in a control group using GGMM

2. To examine specific and generalised effects of training. Specifically, to examine verbal memory (VM performance) following training representing more specific effects of training, as well as long-term verbal memory (LTVM) and executive function (EF) performance trajectories, representing more generalised effects.
3. To identify individual baseline characteristics, such as age, sex and proxies for CR (education and estimated premorbid IQ) that predict and therefore distinguish the three different cognitive performance trajectories (*i.e*. the differential class membership for VM, LTVM and EF) for the study sample.

**Hypotheses**

Based on dominant theory and past research, three hypotheses were made:

1. *Inter-individual improvements in cognitive performance following training.* Firstly, it was hypothesised that GGMM analysis would demonstrate inter-individual differences in training effects with some participants undertaking the training demonstrating significantly improved cognitive performance trajectories compared to controls.
2. *Specific and generalised training effects.*

   Secondly it was hypothesised that both specific (VM) and generalised training effects (LTVM and EF) of training on cognitive performance would be identified in cognitive performance trajectories of the treatment group compared to controls.

3. *Baseline characteristics as predictors of class membership.* Thirdly, it was hypothesised that baseline characteristics – namely age, sex, and proxies for CR, education and estimated premorbid IQ – would predict individual membership of the three inter-individual cognitive performance trajectories (VM, LTVM, and EF). In particular, it was predicted that individuals who were younger, had a higher estimated premorbid IQ and education would be represented in the highest performing class (*i.e.* those showing the greatest trajectories of improvement across the 12-month follow-up period). Females were also predicted to show the greatest improvements, despite the evidence in the literature being considerably mixed with regard to the effects of sex on cognitive performance (Herlitz et al., 1997; Lezak et al., 2004; Zelinski, Gilewski, & Schaie, 1993).

## Overview of the Thesis

Chapters 2 to 5 consist of a literature review and Chapters 6 to 10 are empirical chapters.

*Chapter 2* covers the profile of ARCD from a neuropsychological perspective and importantly highlights the emerging empirical evidence that inter-individual differences exist amongst older adults. A definition of 'successful ageing' is also provided.

*Chapter 3* explores theories explaining heterogeneity of manifestations of ARCD and successful ageing: CR, plasticity (neuroplasticity and cognitive plasticity) and the 'Use it or lose it' hypothesis. It highlights the combination of inherent mechanisms and importantly the impact of environmental influences. Supportive evidence from epidemiological, cross-sectional and longitudinal studies in older adults is presented.

It also highlights controversies surrounding current studies and investigates these, specifically noting experimental design deficits and problems with existing statistical approaches.

*Chapter 4* discusses specific cognitive training paradigms, considered to represent the direct experimental application of an enriched environment, thereby testing the effects of the 'Use it or lose it' and associated theories. It defines cognitive training and discusses key areas requiring investigation within the literature, including cognitive training studies, to best interpret outcomes. These key points include exploring the efficacy of training paradigms from a group perspective and the legitimacy of these studies, including discussion surrounding the specificity, generalisability and long-term effects of training. It also highlights the need to consider and account for inter-individual differences in training responsiveness. Chapter 4 further details the significant literature looking at individual baseline characteristics, including age and sex. It discusses proxies for CR – education and indices of intelligence – that have been investigated as being associated with, or predictive of ARCD. It also explores the factors that may help explain heterogeneity of cognitive training outcomes and assist in identifying who benefits from cognitive training. Explanations of the possible mechanisms at work influencing the interplay between such predictors and cognitive training are also briefly discussed.

*Chapter 5,* the final chapter of the literature review, highlights GBGM techniques, including GGMM. The chapter outlines and discusses the statistical standing of this new type of modelling technique, from a largely non-technical perspective. The chapter also highlights the key empirical components, and the common and recommended analytic procedures by which the models are produced and selected. In addition, limitations to these techniques are elucidated. To further explain the concepts discussed, a worked example based on the current study is used throughout the chapter.

*Chapters 6–10* comprise details of the current cognitive training intervention study. Chapter 6 describes the methodology used. Chapter 7 presents the VM results. Following standard reporting procedures, results are shown separately for the control, experimental and from the entire cohort (the joint group). Also presented are the GGMM results, applied to the joint data and used to identify and predict VM performance trajectories following training. Chapters 8 and 9 present results following the same format for the more distal domains from training: the LTVM and EF results, respectively.

*Chapter 10* offers a discussion of the results of each of the cognitive domains: VM, LTVM and EF. It gives an overall summary of the main findings from the three cognitive domains as well as the statistical and theoretical implications, limitations, suggested future directions for research and the final conclusions drawn.

# Chapter 2

## Age-related Cognitive Decline: Population-level and Inter-individual Profiles

The ageing literature generally demonstrates that from a population perspective, average levels of cognitive performance in many cognitive domains gradually decline with age. Recently, however, striking inter-individual differences in the manifestation of this change have been noted (Buitenweg, Murre & Ridderinkhof., 2012; Jolles & Crone, 2012; Karbach & Schubert, 2013; Leoutsakos, Muthén, Breitner, & Lyketsos, 2012; Martin et al., 2011; Park, 2007; Terrera et al., 2010; West & Hastings, 2011). Accordingly, investigations into age-related cognitive decline (ARCD) have begun to shift away from a largely population-level approach towards considering the importance of heterogeneity in the manifestations of cognitive decline.

As highlighted in the Introduction chapter, there are several core questions that are essential to the exploration of this burgeoning area of enquiry and are consequently the focus of this study. These include what theories and mechanisms explain these differences, and what can be done to improve cognitive performance in older adults. In addition, it is useful to explore what predictors are used to explain responsiveness to interventions designed to improve cognitive function. Prior to the comprehensive investigation of these questions presented in this thesis, it is useful to describe and define ageing from a cognitive perspective. Evidence of heterogeneous manifestations of ARCD are outlined in this literature review and appropriate statistical methods to investigate inter-individual differences in rates of decline are also elucidated.

## Definition of Cognitive Ageing and Age-related Cognitive Decline

### Cognitive Ageing

Cognitive ageing is considered a process in which individuals demonstrate robust, progressive deficits in cognitive performance (Buckner, 2004; Deary, 2009; Hedden & Gabrieli, 2004; Schaie, 2000). The ageing literature is littered with both descriptive terms and formal sets of criteria to describe the widely-reported, population-level, age-related loss of cognitive function many of which are broadly similar. None of these terms and criteria are universally accepted or applied. Classifications include Benign Senescent Forgetfulness (BSF; Kral, 1962), Age-associated Memory Impairment (AAMI; Crook, Kisvarday & Eysel, 1998), Age-associated Cognitive Decline (AACD; Levy, 1994) and Cognitive Impairment, No Dementia (CIND; Graham et al., 1997).

This thesis will follow the Diagnostic and Statistical Manual, 4th Edition, Text Revision (DSM-IV-TR; American Psychiatric Association [APA], 2000) terminology and refers to cognitive decline with ageing as *Age-related Cognitive Decline* (ARCD). This definition is used when the decline is considered within normal (non-pathological), age limits and has an objective, meaningful impact on day-to-day cognitive function[2]. It is for this reason the current review uses the term ARCD.

Another common term in the literature, MCI is considered a separate construct (also with multiple definitions based on, for example biomarkers and cognitive

[2] This will be further elaborated below.

performance) and the newly minted Mild Neurocognitive Disorder (DSM-V, APA-, 2013). Underlying these definitions is, however, an implicit pathological reason for the decline. As such, literature focusing on MCI is excluded from the present thesis.

**Profile of Age-related Cognitive Decline from a Population-Level**

Age-related cognitive decline considered at a population-level (*i.e.*, based on cross-sectional and longitudinal comparisons) consistently reveal that increases in age are associated with lower levels of performance on a wide range of cognitive measures, including declines in verbal memory (VM), processing speed, visuo-spatial skills, attention and executive functions, including working memory (Buckner, 2004; Deary et al., 2009a; Hedden & Gabrieli, 2004; Schaie, 2000). At a population level, this decline has been shown to begin from middle age onwards, although some studies suggest performance in some domains begins to weaken in younger adulthood, *e.g.*, individuals in their 30s and 40s (Hedden & Gabrieli, 2004; Park & Reuter-Lorenz, 2009). Cognitive deterioration can be quite substantial. For example, speed of processing and declarative memory performance has been shown to drop almost two standard deviations (*SDs*) across the adult lifespan (Der & Deary, 2006; Kramer & Willis, 2002; Park et al., 2002; Salthouse, 1996; Schaie, 1994; Schaie, 1996; Verhaeghen & Salthouse, 1997).

Although ARCD does not profoundly affect real-world function, as noted above, all cognitive domains affected are considered important for carrying out everyday activities and other cognitively demanding tasks, living independently and leading an engaged and fulfilling life. Even modest reductions in cognitive function in older adults can engender functional deficits and can negatively impact on quality of life, independence and performance of instrumental activities of daily living. It also impacts on the frequency and quality of social interaction and engagement in cognitively stimulating activities (Boron et al., 2007a; Deary et al., 2009a; Lawton, 1982; Mahncke et al., 2006a). More specifically, ARCD can manifest as problems beyond remembering names and appointments (mentioned earlier), such as medical adherence and other health maintenance behaviours and/or difficulties in solving complex problems, including financial management (APA, 2000; Park et al., 2007). There is also a link between declining cognition and physical decline (Black & Rush, 2002). Functional decline leads to increased occurrence of disability and morbidity and further lessens independence and quality of life in older adults.

As previously highlighted, age-related cognitive decline can be best understood as representing a 'normal' process or 'subclinical' cognitive decline, consistent with a person's age and education level. That is, the dominant conceptualisation suggests ARCD is not the result of pathological processes,[3] such as a mental disorder or a neurological condition including dementia processes, nor mild cognitive impairment (MCI[4]; DSM-IV-TR; APA, 2000; Storandt, 2008). That said, it is often still considered on a continuum of cognitive decline with MCI and dementia, whereby a clinical diagnosis is rendered when the individual reaches a critical threshold of dysfunction in their cognitive performance trajectory across time (Tucker-Drob & Salthouse, 2011).

### Profile of Age-related Cognitive Decline on an Inter-individual Level

As previously noted, one of the most striking aspects of many investigations into ARCD is large individual differences with regard to the expression of ARCD (Barnes

[3] Whilst remains contentious issue (c.f. Wilson et al., 2010) this is the view taken in the present particularly given its distinct cognitive presentation from dementia and MCI.
[4] MCI also has multiple difinitions, e.g. NIA-AA

et al., 2007; Lindenberger & von Oertzen, 2006; Nelson & Dannefer, 1992; Schaie, 1994). Some older adults maintain high levels of cognitive performance while others experience a rapid decline (Tucker-Drob & Salthouse, 2011). Indeed, the significant cognitive variability across individuals has been considered by some researchers as a *sine qua non* of later life development (Raz, 2009). While many studies explore ARCD simply from a group-level, it is inter-individual differences that have significant implications for the accuracy of common-place statistical investigations (Baltes, Freund, & Li, 2005; Kramer & Willis, 2002). This raises two important issues: whether it is appropriate to generalise group data in relation to cognitive function in ageing; and, if there are defining characteristics of those individuals who withstand the effects of ageing on cognition. Chapter 5 includes further discussion into ideal, powerful statistical approaches – group-based growth modelling (GBGM; *e.g.*, Muthén & Shedden, 1999; Nagin, 1999) – which move away from group-level statistics. Such a statistical approach is considered more appropriate when examining cognitive changes in heterogeneous groups of older adults. Indeed the associated empirical study specifically addresses the need for identifying heterogeneity in cognitive function from an individual perspective, over-time (i.e. not impairment in performance relative to norms or controls as is the case for cross sectional studies). For now, however, the characteristics of individuals who appear to be withstanding the effects of ageing on cognition are considered. These individuals are referred to as exhibiting *successful ageing.*

**Successful Ageing**

The ageing literature is replete with different terms describing successful ageing. Most definitions incorporate multiple health domains and extend beyond the maintenance of cognitive health. Other definitions include sustained physical health, social health and high ratings of quality of life (APA, 2000; *cf.* definitions by Butler, 1991; Rowe & Kahn, 1987; World Health Organization [WHO], 2002). The current thesis and accompanying empirical study will nonetheless use the term *successful ageing in relation to cognitive function*.[5] Whilst all factors associated with successful ageing are considered important, the present review will focus on some of the possible factors influencing cognitive health. In the cognitive context, successful ageing is considered multi-dimensional and lacks a universal operationalised definition. It has been variously conceptualised in an ipsative, normative or criterion-referenced manner (Depp, Harmell, & Vahia, 2012). For example, some definitions exclude clinical impairment such as dementia, yet fail to identify factors that determine optimal cognitive ageing (Habib, Nyberg, Nilsson, 2007; Yaffe et al., 2009). Other researchers specify successful ageing as being exhibited by individuals who maintain their peak cognitive performance level during senescence, with few or no decrements in cognitive functioning in domains typically seen in epidemiological, cross-sectional and longitudinal studies of decline (Barnes et al., 2007; Carey, 2007; Jones et al., 2005; Lindenberger & von Oertzen, 2006; Nelson & Dannefer, 1992; Raz, 2009; Schaie, 1994; Yaffe et al., 2009). "Successful agers" have also been described as those tending to continue cognitive development later than most; they typically reach their cognitive asymptotes in late midlife. They are also likely to maintain their overall level of cognitive functioning until shortly before death (Schaie, 2008). Overall, successful ageing is experienced by only a minority of individuals (Carey, 2007).

[5] The cognitive component of successful ageing is also referred to in the literature as cognitive vitality (Kramer et al., 2004).

The working definition of successful ageing in the current thesis combines many of the above-mentioned cognitive factors. Thus successful ageing is considered to be experienced by a minority of individuals who exhibit few or no decrements in cognitive functioning in domains typically seen to decline in group-level analyses such as ARCD (Barnes et al., 2007; Carey, 2007; Jones et al., 2005; Lindenberger & von Oertzen, 2006; Nelson & Dannefer, 1992; Raz, 2009; Schaie, 1994; Yaffe et al., 2009). Thus for the purposes of this review ARCD and successful ageing are distinct.

**The Expression of Inter-individual Differences**

A number of longitudinal investigations into ARCD have provided reliable evidence of individual differences, including amongst those who demonstrate successful ageing, particularly in the memory domain (Christensen et al., 1999; deFrias, Lövdén, Lindenberger, & Nilsson, 2007; Hultsch, Dixon, & Small, 1998; Lindenberger & Ghisletta, 2009; Lövdén et al., 2004). More specifically, inter-individual differences in ARCD have been described as varying not only in relation to both the timing but, importantly, in the rate of decline across individuals (Macdonald, Hultsch, & Dixon, 2003; Schaie, 1996).

**Inter-individual differences in the commencement of age-related cognitive decline.** When looking at the timing of decline, despite group-data predictions, observed changes in cognition generally are not a direct reflection of age and age-graded biological limits. For example, some individuals retain high cognitive functioning into their eighth and ninth decades, whilst others show severe cognitive impairment early in the ageing process (Hedden & Gabrieli, 2004; Hertzog, Kramer, Wilson & Lindenberger, 2008a). Differences across individuals may, however, increase at older ages (deFrias, 2007; Fandakova, et al., 2012; Kliegl et al., 1989;

Kliegl et al., 1990). deFrias et al. (2007) demonstrated that whilst there were mean decreases in changes in episodic memory recall performance, inter-individual differences tend to increase with age. Increased variability of performance has also been demonstrated in regression analyses (Hertzog et al., 2009; Kliegl et al., 1989; Kliegl et al., 1990). Older individuals may still have the capability to move up or down within this range, potentially as a result of environmental influences, including specific cognitive training (Denney, 1984). This is further discussed in Chapter 4.

It should be noted, however, that evidence for increased inter-individual variability with age is mixed in cross-sectional studies (*e.g.*, Morse, 1993; Lindenberger & Baltes, 1997). This may be a consequence of the limitations of cross-sectional approaches, such as the influence of selectivity and failures of the assumption of interval measurement in cognitive tests (Tucker-Drob & Salthouse, 2011). As longitudinal studies enable clearer inferences to be drawn about individual differences in the rates of cognitive change, and given that the following empirical study utilises this method, the present review will mainly focus on the evidence from such studies.

**Inter-individual differences in the rates of age-related cognitive decline.** Recently longitudinal studies are beginning to implement statistical approaches allowing consideration of inter-individual differences in trajectories of change (Nagin, 1999; Muthén, 2004). Regression models have frequently been used to estimate rates of change over time (*i.e.*, slopes). More sophisticated regression models – such as the latent growth modelling (LGM) techniques – enable researchers to estimate variations in changes in cognitive ageing trajectories that are, in theory, error free. For example, LGM, similar to (although not identical with) multilevel, random effects modelling (Jung & Wickrama, 2008; Laird & Ware, 1982; Rao, 1958; Scher, Young, & Meredith, 1960; Tucker, 1958) is an example of this improved approach. Such models account for both performance across time and variations across individuals. In addition to measuring the group (average) rate of performance, each individual's distinct performance, which may be quite different from what happens to the group, is represented by the variation (or random effects) around the slope of the trajectory (Li & Acock, 1999). Variation around the intercept (the initial level) can also be included in analyses (Li & Acock, 1999). Thus, individual differences in growth may be examined by testing for systematic individual differences from a group-level growth model (Reynolds, Gatz, & Pederson, 2002). Latent growth modelling, therefore, provides a better approach than traditional statistical methods in which simple difference scores are utilised and variation is likely to be disproportionately attributed to measurement error (Cronbach & Furby, 1970). Further development of even more sophisticated models can create multiple trajectories to represent the inter-individual trajectories (around which variation can also be measured). These group-based growth models, as previously noted, are elaborated in Chapter 5.

Given the growing evidence for the recognition of systematic and statistically significant variations in cognitive ageing and the development of sophisticated statistical approaches to measure trajectories of change, researchers are beginning to implement growth modelling approaches to determine the temporal manifestation of inter-individual differences in ARCD using longitudinal studies (*e.g.*, Barnes et al., 2007; Tucker-Drob & Salthouse, 2011; Wilson et al., 2002a).

Barnes and colleagues (2007) conducted a longitudinal study of 9,704 community-dwelling older women (mean age at baseline 72 years). They implemented a random-effects regression on performance on the modified MMSE (3MS; Teng & Chui, 1987) over a 15-year period. They reported three distinct cognitive performance trajectories. It was observed that 33% of the cohort had major declines in cognitive functioning, 58% had minor declines (considered to be a normal trajectory), and 9% of the cohort exhibited no decline of global functioning over the 15-year period.

Yaffe and colleagues' (2009) conducted a study of 2,509 well-functioning older males and females who were part of the "Health, Ageing and Body Composition" (Health ABC) study. Participants were aged 70–79 years at recruitment. The study tracked cognitive trajectories across eight years based on performance on the 3MS. Of these participants, 411 (16%) were identified as being major decliners with significant clinical decline, 1,340 (53%) were minor decliners with more modest or "typical age-related" decline (p. 2034), and 758 (30%) were maintainers who exhibited no decline on global cognitive function. Whilst the 3MS is an imperfect measure of cognitive function and ARCD, these studies highlight the importance of considering heterogeneity of cognitive performance across older adults.

Finally, Schaie and Hofer (2001) reviewed 27 ongoing longitudinal studies of psychological ageing up until death (the first of which commenced in 1956). Whilst the authors also importantly highlighted inter-individual differences in cognitive trajectories, they suggested that, unlike the studies described above, four groups of individuals' cognitive trajectories are evident; 1) a small subgroup of individuals who showed very modest decline on highly speeded tasks and were likely to maintain their overall level of cognitive functioning until shortly before their death, these individuals were considered to be exhibiting successful ageing (Fillit et al., 2002; Rowe & Kahn, 1987); 2) individuals considered to be ageing normally, that is, they were demonstrating a common trajectory of ARCD decline, mentioned earlier; 3) individuals who went on to develop MCI; and, 4) individuals who could be diagnosed with dementia.

The above studies clearly demonstrate that inter-individual differences in cognitive trajectories can be shown for those exhibiting ARCD type profiles, for those exhibiting successful ageing and also for those at the opposite end of the spectrum, that is, individuals' expressing levels of cognitive function allowing for dementia diagnoses. Indeed these studies also pave the way for further growth modelling research, which needs to balance both statistical and theoretical considerations (this concept will be further discussed in chaper 5).

**Summary**

In summary, at a group-level ARCD is commonly associated with decline in a number of abilities that are important for carrying out everyday activities and leading an engaged and fulfilling life. There are, however, striking inter-individual differences in the manifestations of ARCD. A minority of individuals experiencing successful ageing; individuals exhibiting little to no large decrements in cognitive functioning in domains typically seen in epidemiological, cross-sectional and longitudinal studies., Ascertaining the exact timing and trajectory of ARCD is highly influenced by inter-individual differences and the statistical analyses used when investigating change. In order to measure true cognitive changes, the application of appropriate statistical techniques is required, such as GBGM.

As noted at the beginning of the chapter, there are several core questions that are essential to the exploration of inter-individual differences, including explanatory theories and the purported mechanisms underlying those theories. Chapter 3 will further elaborate on these neuropsychological and neurobiological constructs and what is believed to contribute to successful ageing. The chapter will discuss the evidence for and controversies surrounding epidemiological, cross-sectional and longitudinal animal and human imaging studies. Whilst investigating these theories, the discussion will identify further design deficits and problems that are intrinsic to the existing statistical approaches implemented thus far.

**LITERATURE REVIEW**

# Chapter 3

## Theories Explaining Heterogeneity of Manifestations of Age-related Cognitive Decline and Successful Ageing

The emerging evidence of heterogeneity in individual trajectories of cognitive ageing, including those deemed to be 'successfully ageing' versus those exhibiting age-related cognitive decline (ARCD), has led many researchers to investigate how we can mimic individuals' successful ageing. This area is of major importance to public health and to older individuals themselves, as noted in Chapter 1. Indeed, the term successful ageing was introduced to explore the idea that age-related decline may be avoided (Yaffe et al., 2009).

There have been multiple lines of research examining potential causes of heterogeneity in older adults' cognitive performances. This research encompasses many scientific fields, including genetics, epigenetics, neurobiology, brain activation, general health, chronic illness, psychological and personality characteristics and lifestyle factors (Anstey & Christensen, 2000; Greenwood & Parasuraman, 2010; Johnson, Bengsteon, Coleman, & Kirkwood, 2005; Schaie, 2008; Rabbitt et al., 2004). The present review takes a biopsychosocial focus and explores three of the most prominent theories in the cognitive psychology literature. Specifically, the theories discussed include the concept of cognitive reserve (CR), prominent in neuropsychological literature, and the related neurobiological concept of plasticity. Cognitive reserve plasticity will also be discussed in relation to a third, central overarching and predominating hypothesis, popularly known 'Use it or lose it'. These theories are heuristically useful in examining and understanding individuals who age successfully.

These three similar, linked and overlapping theories maintain an emphasis on the two key factors in relation to an individual's expression (or lack thereof) of ARCD: 1) an individuals' inherent capacity, and 2) the potentially enriching effects of environmental experience. In particular, intrinsic and environmental influences across the normal lifespan have been suggested to have an effect on each individuals' brain and cognitive functioning (Christensen et al., 1999; Bialystok et al., 2004; Lindenberger et al., 2008), thus exaggerating inter-individual variability as individuals grow older (Buitenweg et al., 2012). Most disciplines recognise the multifactorial nature of successful ageing, including the influence (and interaction) of several factors.

This chapter discusses major avenues of evidence relating to the three theories, stemming from a large body of epidemiological, cross-sectional and longitudinal studies, as well as imaging studies. Supportive studies will mostly be described and discussed in relation to all three concepts, given the large degree of interchangeability of the evidence used to account for each theory across the literature.

## Cognitive Reserve

Cognitive reserve was originally introduced to explain the imperfect coupling between the degree of pathological biomarkers in the brain with cognitive function. Two classic studies (Roth, Tomlinson, & Blessed, 1966; Tomlinson, Blessed & Roth (1976) demonstrated that some individuals with an extensive distribution of amyloid ('senile') plaques at autopsy (*i.e.*, at levels required to meet neuropathologic criteria for Alzheimer's disease [AD]) did not exhibit pre-death cognitive impairment. Indeed, it has been estimated that up to 30% of individuals with moderate to severe levels of neurodegenerative pathology at autopsy show no signs of cognitive dysfunction at ante-mortem test (Neuropathology Group of the, Medical Research Council Cognitive Function and Ageing Study [MRC CFAS], 2001). In these cases, CR is considered to provide a general protective function, such that cognitive performance can be maintained despite pathological damage to the brain.

This lack of concordance between brain damage and cognitive performance has since been extended to the 'normal' ageing context, whereby individuals with a greater CR are predicted to be able to compensate for age-related brain changes before demonstrating manifestations of those changes (Buiza et al., 2008; Buckner et al., 2004; Fratiglioni, Paillard-Borg, & Winblad, 2004; Gatz, 2005; Stern, 2002; Stern et al., 2003; Stern, 2007; Willis, Schaie, & Martin, 2009). Thus, in this case, the CR hypothesis posits that this phenomenon could be explained by individual differences in the ability to cope with age-related disruptions to brain function (Stern et al., 2003). Cognitive reserve has therefore been explored in relation to cognitive decline in older adults (*e.g.*, Cervilla, Prince, Lovestone, Mann, & Joels, 2002; Christensen et al., 1997; Kramer et al., 2004).

**Passive and Active Models of Reserve**

Cognitive reserve can be conceptualised using both passive and active models. Stern's (2003) passive threshold model, known as "brain reserve capacity" (BRC; p.589), relates to the physical capacity of the brain and is also referred to as a passive or hardware model (Scarmeas & Stern, 2003). Brain reserve capacity has been measured in a variety of ways, including neural substrates corresponding to reserve, such as the number of large pyramidal neurons, intra-cranial volume (ICV), as well as head girth and circumference (Barulli & Stern, 2013; Gatz, 2005; Stern, 2002). These structures (*i.e.*, passive components) are thought to add capacity to efficient processing of information, enhanced retrieval of memories and problem solving (Stern et al., 2003; Whalley, Deary, Appleton, & Starr, 2004). Thus, a greater BRC would also enable greater tolerance to the same, age-related brain changes, than a smaller BRC, before reaching the threshold for exhibiting ARCD. This concept is supported by a notable proportion of autopsy cases (Mortimer, 1997; Valenzuela & Sachdev, 2006a).

In contrast, the 'active' (or software) model of reserve focuses on the manner in which tasks are processed. This is the more commonly emphasised theory in the context of the current thesis. At a cognitive level, it is considered that there are typically multiple pathways by which effective brain function can be achieved (Hunt, 1978; Lautrey, 2003). The active model is used to highlight the capacity to optimise or maximise performance through the differential recruitment of *new* brain networks or cognitive pathways (neural compensation; Stern, 2009). Alternatively, or in addition to the use of new neural pathways, the theory can be used to argue that as a result of CR there may be *more efficient or flexible* ways to use *existing* brain networks and cognitive strategies (neural reserve; Gatz, 2005; Stern, 2002; Stern, 2009; Willis et al., 2009) despite the neurobiological processes contributing to ARCD. Thus, whilst strictly speaking the mechanism by which active reserve manifests is anatomical, it is still considered an active process concerning the manner in which the brain functions (Whalley et al., 2004).

## Cognitive Reserve and Successful Agers

**Inherent capacity.** When considering CR from the perspective of inter-individual differences, successful agers can better shift or develop new strategies if the negative physiological effects of age or the individual's contextual (environmental) demands require them to do so (Hertzog & Dunlosky, 2004; Hertzog & Robinson, 2005; Schunn & Reder, 2001; Touron & Hertzog, 2004). This may be a reflection of automatic and/or conscious processes for more efficient processing. In relation to the conscious processes, for example, Noack and colleagues (2009) consider the role of higher-order functioning in which the successfully ageing individual better monitors their current cognitive state, performance outcomes and new learning. This enables these individuals to achieve better cognitive performance. In contrast, an individual demonstrating a greater level of ARCD may have a compromised or underutilised capacity for such active (conscious) cognitive changes.

Thus, overall, with the utilisation of new or pre-existing neural networks via automatic and/or conscious choice, this component of CR can represent an inherent capacity within a successfully ageing individual (Hertzog & Robinson, 2005; Nelson & Narens, 1990; Stine-Morrow, Miller & Hertzog, 2006).

**Enriching cognitive environment**. The second key component of CR theory considers the prediction that those who had more enriching cognitive environments – from childhood and throughout their lifespan – have more resilient neurobiological and/or cognitive architectures. The enriching experiences are purported to have built CR to protect against the expression of ARCD (Tucker-Drob & Salthouse, 2011). Examples of enriching experiences include education, stimulating occupations and other lifestyle factors, such as social engagement, physical activity and mentally stimulating leisure activities. It has also been suggested that various combinations of these factors, discussed as building CR, do so either through separate or synergistic effects (Staff et al., 2004; Stern et al., 2003; Stern, 2009; Valenzuela, 2008).

Of key significance to this discussion is that CR is at least partly malleable through environmental manipulations and importantly it offers an avenue through which older adults could become more resilient to ARCD. Furthermore, this may be purposefully manipulated through interventions so that older adults can be assisted to strengthen their neural and cognitive architectures. They can be helped to become more capable of making the necessary efforts directed at selecting and implementing strategies that are appropriate for a greater variety of tasks (Carey, 2007; Jones et al., 2006; Noack, Lövdén, Schmiedek, & Lindenberger, 2090). This is discussed in the context of cognitive training in Chapter 4.

**Proxies for Cognitive Reserve**

Given that CR is a hypothetical construct it cannot be measured directly. Proxies for CR are therefore used, and take into account both inherent capacity and experience. Indices of intelligence are used to reflect the intrinsic component, including measures of crystallised intelligences – such as verbal knowledge – IQ and estimates of premorbid IQ. It has been suggested that in some cases premorbid IQ alone might be a more powerful measure of reserve than alternatives (Albert & Teresi, 1999; Alexander et al., 1997; Stern, 2002; Valenzuela & Sachdev, 2006a; Valenzuela & Sachdev, 2006b). Indices of intelligence have been explored in relation to cognitive decline in older adults (*e.g.*, Cervilla et al., 2002; Christensen et al., 1997; Kramer et al., 2004).

Education can also represent the environmental contributors to CR, determined by external experiences, and the most popular non-biological proxy (Albert & Teresi, 1999; Alexander et al., 1997; Goldin, 1998; Jones et al., 2011; Reynolds et al., 2002; Stern, 2002; Schaie, Willis, & Pennak, 2005; Treiber et al., 2011; Tucker-Drob &

Salthouse, 2011; Valenzuela & Sachdev, 2006a; Valenzuela & Sachdev, 2006b; Yaffe et al., 2009). See Jones and colleagues (2011) for a critique of proxy measures for CR.

## Plasticity

Another related concept discussed in the ageing literature is plasticity. The term *plasticity* has been used for over a century; however its meaning has changed significantly over time and currently has multiple connotations across different disciplines (Lövdén et al., 2010).

Whilst the term *plastic* refers to changeability, malleability and modifiability, a common contemporary interpretation of plasticity relates to the capacity for change. The current scientific conceptualisation highlights that this capacity is a reactive phenomenon, reflecting secondary change in response to an initial alteration in the system (Lövdén et al., 2010). It emphasises the dynamic nature of the brain, which has a number of different and interacting levels. Whilst plasticity is most often discussed as expressing positive change, this is not always the case (Hertzog, Price & Dunlosky, 2008b). Thus either positive or negative changes are instigated.

As highlighted in CR theory, these changes may be a result of inherent mechanisms and/or the environment. The plasticity literature in animal models emphasises *environmental enrichment* (EE), incorporating complex, novel environments. This has been extrapolated to humans using the terms *cognitive enrichment (or* the *arousal hypothesis).* Here cognitive enrichment involves mental stimulation within sociocultural contexts and personal behaviours, *e.g.*, lifestyle pursuits, education (Greenwood, 2007; Baltes, Reuter-Lorenz, & Rösler, 2006b; Willis, Schaie, & Martin, 2009). Positive change may result from normal brain development or learning and memory (Olesen, Westerberg, & Klingberg, 2004; Willis et al., 2009). These changes create further intrinsically-generated structural, functional, cognitive and behavioural responses to the environmental stimuli (Greenwood, 2007). The adult brain is capable of positive plasticity, even into older age, and thus is highly pertinent to this discussion of successful ageing and ARCD. Plasticity mechanisms with positive outcomes have been shown to slow or reverse brain ageing in a number of animal studies through evidence of enhanced mental activity (Nithianantharajah & Hannan, 2006). This is described in the literature as *functional plasticity of ageing* (Greenwood, 2007).

Possible negative changes include the ageing process itself, certain pathological conditions, injurious states (*e.g.*, brain damage or epilepsy) and sensory deprivation (Kempermann, Gast, & Gage, 2002; Noack et al., 2009; Whalley, 2004).

This thesis considers plasticity with regard to the dynamic interplay between the ageing process and the enriching environment. The emphasis in the discussion below is on the expression of these changes, in this case cognitive changes in domains associated with ARCD.

When plasticity is discussed from the neurobiological perspective it encompasses neuroplasticity and cognitive plasticity, concepts which are tightly linked (Barulli & Stern, 2013). A similar construct, cognitive flexibility is also discussed and is often distinguished in relation to cognitive plasticity (Lövdén et al., 2010).

**Neuroplasticity**

Neuroplasticity is also called neural, cortical, cerebral or brain plasticity and refers to the physiological capacity of an individual to make neural changes. Neuroscientific evidence (*e.g.*, results from genetic and histological animal studies) provides detailed information about neuroplasticity mechanisms (*cf*. Anderson et al., 2001; Bhardwaj et al., 2006; Carey, 2007; Ehninger, & Kempermann, 2003; Gould & Gross, 2002; Greenwood, 2007; Hertzog et al., 2009; Krech et al., 1956; Mattson et al., 2002; Rakic, 2002; Rosenzweig et al., 1962; Trachtenberg et al., 2002). The neural basis of these changes in the human brain have been explored indirectly using positron emission tomography (PET) and more recently functional magnetic resonance imaging (fMRI) studies through monitoring increases in blood flow and oxygenation in specific brain structures (Cabeza, Anderson, Locantore, & McIntosh, 2002; Lövdén et al., 2010; Park & Reuter-Lorenz, 2009). These occur in response to fluctuations in neural or glial activity and are represented by reorganisation of neural circuits, supporting neurological architecture and changes in function (Baltes, Lindenberger, & Staudinger, 2006a). The theory of positive neuroplasticity has also been purported as exemplifying the mechanisms by which CR builds (Kramer et al., 2004; Stern, 2007; Wilson et al., 2009).[6]

**Cognitive Plasticity**

Cognitive plasticity is a multifaceted concept that is considered to be modifiable at all phases of development, including older age, like neuroplasticity (Baltes et al., 2006a). Similar to neuroplasticity and CR, cognitive plasticity involves a consideration of both inherent capacity and environmental influences (Baltes & Lindenberger, 1988; Willis et al., 2009). It is often used to denote improvements in the acquisition or building of cognitive function to improve cognitive performance (*e.g.*, knowledge, memory, processing speed and efficiency; Boyke et al., 2008; Baltes & Lindenberger, 1988; Draganski et al., 2004; Draganski et al., 2006; Willis et al., 2009). This is

[6] Stern et al.'s (2003) model of CR suggests that passive CR (*e.g.*, an inherently bigger brain) can exist even when neuroplasticity has been compromised.

discussed extensively in relation to 'Use it or lose it' and successful ageing below, and in relation to cognitive training in older adults in Chapter 4.

The link between cognitive plasticity and neuroplasticity is also highlighted when considering the mechanisms by which cognitive plasticity builds. Manifestations of cognitive plasticity depend upon neuroplasticity (Greenwood & Parasuraman, 2010). Indeed, experimental animal studies indicate that alterations in neurons, synapses, glial cells, *etc.*, result in changes in cognitive plasticity as evidenced by improved cognitive performance (Baltes & Lindenberger, 1988; Baltes, Staudinger & Lindenberger, 1999; Cabeza, 2002; Colcombe, Kramer, Erickson, & Scalf, 2005; Kramer & Willis, 2002; Park & Reuter-Lorenz, 2009; Reuter-Lorenz & Cappel, 2008; Salthouse, 1984; Willis et al., 2009). That said, the directionality of the association has not been fully elucidated, and cognitive shifts may drive neuroplasticity via neural competition and rearrangements (Baltes & Lindenberger, 1988; Baltes et al., 1999; Cabeza, 2002; Colcombe et al, 2005; Greenwood & Parasuraman, 2010; Kramer & Willis, 2002; Park & Reuter-Lorenz, 2009; Reuter-Lorenz & Cappel, 2008; Salthouse, 1984; Willis et al., 2009).

**Cognitive flexibility.** It should also be noted that the structural change associated with cognitive plasticity is considered in contrast to *cognitive flexibility* by some researchers (Fandakova et al., 2012; Lövdén et al., 2010; Noack et al., 2009). As mentioned earlier in the context of CR, cognitive flexibility refers to the utilisation of existing (*i.e.*, previously formed) neural pathways, thereby falling within an individuals' pre-existing range of cognitive capacity, known as their functional supply (Lövdén et al., 2010). Cognitive flexibility also refers to the adaptation of a pre-existing behavioural repertoire, for example specific conscious application of strategies to execute a compensatory goal-directed action (Carey, 2007; Fandakova et al., 2012; Hill, Backman, & Stigsdotter-Neely, 2000; Jones et al., 2006; Lövdén et al., 2010; Lövdén et al., 2012; Noack et al., 2009).

In contrast, cognitive plasticity refers to the expansion of this repertoire following structural change which occurs at an unconscious or automatic level (Lövdén et al., 2012). That is, flexibility-based changes use existing cognitive capacity whereas plasticity-based changes result in alterations in cognitive capacity (Lövdén et al., 2010). For example, when an individual takes up a new hobby, he or she may engage in cognitive plasticity resulting in the creation of new neural networks and would thereby have developed a new skill set. In contrast, an individual previously practising that hobby who wants to increase his or her skill set through more practice is using their brain more flexibly.

**Positive Plasticity Through Training**

In summary, CR and plasticity theories are tightly linked constructs. Both highlight two key aspects of cognitive performance in older adults: an individual's inherent cognitive capacity and the effect of specific enriched contextual conditions. Plasticity in particular highlights the biological and cognitive underpinnings in relation to experience. Importantly both theories imply that it may be possible for older individuals enhance their cognitive potential via purposeful positive manipulation of their environment. Cognitive reserve and plasticity theories are discussed further below in the context of the 'Use it or lose it' hypothesis.

## 'Use It or Lose It'

The popularly-termed 'Use it or lose it' hypothesis'[7] has numerous specific nuances depending on the area of study. Ultimately the hypothesis encapsulates, like both CR and plasticity theories, that 'Use' of one's cognitive systems – by performing stimulating and demanding activities in an enriched environment – could lead to preserved and/or improved cognitive functioning in older adults. Specifically the hypothesis highlights that factors such as an intellectually varied lifestyle and experiences can potentially enhance cognitive performance and promote successful cognitive ageing (Deary et al., 2009a; Hertzog et al., 2009; Valenzuela & Sachdev, 2009).

Furthermore, the 'Use it or lose it' hypothesis underscores that, without cognitive enrichment, successful cognitive performance may be 'lost' (*e.g.*, through comparatively sedentary lifestyles). This is similar to negative plasticity discussed above, as well as Salthouse's disuse perspective (1991) whereby reduced activation of cognitive resources leads to a decrease of cognitive function. Older individuals often change their activity patterns, such that the level of engagement in cognitively demanding activities is lessened (Hultsch, Hertzog, Small, & Dixon, 1999). There is also the suggestion that disuse may occur when tasks become more difficult and there is subsequent disengagement or disuse, *e.g.*, when cognitive decline, health (such as sensory deficits) or logistical obstacles limit the ability to perform such activities (Frick & Benoit, 2010; Hultsch et al., 1999; Scarmeas & Stern, 2003). Changes in activity patterns can also be through specific choice, for example, if an individual has

[7] Similar hyptheses in the literature include the mental exercise hypothesis cognitive and enrichment hypotheses also often incorporates social engagement and physical exercise as an important constituents (Hertzog, Kramer, Wilson, & Lindenberger, 2008a; Katzman, 1993; Salthouse, 1981; Schooler, 1987; Tucker-Drob & Salthouse, 2011; Wilson et al., 2009).

an awareness of signs of ARCD and chooses to disengage for fear of embarrassment, retirement, when an individual 'rests on their laurels', and/or pursues activities in which they already excel (Mahncke et al., 2006a). Changes in behaviour may also reinforce a downward spiral of degraded cognitive function (Mahncke et al., 2006b). Thus, according to the 'Use it or lose it' hypothesis individuals may need to maintain 'Use' of neural and cognitive systems or they will 'lose' cognitive function. In line with the hypothesis, 'successful agers' may 'Use it' through exposure to enriching cognitive environments and/or less cognitive disengagement and therefore demonstrate less cognitive decline.

### Evidence of the Impact of 'Use It or Lose It' Hypothesis, Cognitive Reserve, Plasticity Theories

Research across a number of scientific fields has explored the 'Use it or lose it' hypothesis, and the CR and plasticity theories. Already highlighted within discussions of CR and plasticity theories, there is growing evidence for their contributions to individual differences in normal cognitive ageing and the factors associated with successful ageing (*e.g.*, functional brain imaging).

Epidemiological, cross-sectional and longitudinal studies lend support for the 'Use it or lose it' and associated theories, and are highlighted below.[8] In particular, such studies underscore the concept that environmental influences or cognitive enrichment can cause improvements that are demonstrated even at the group-level.

[8] Most direct effects of cognitive enrichment that have been explored in the cognitive intervention literature specifically investigate applied attempts to mimic enrichment effects (or supplement the benefits of lifetime enrichment; Hertzog et al., 2009; Papp et al., 2009). This is discussed in Chapter 4.

### Cognitive Enrichment in Epidemiological, Cross-sectional and Correlational Studies

There are numerous epidemiological, cross-sectional and correlational studies demonstrating an association between the level of participation in intellectual, social and physical activities and performance on various cognitive tasks in healthy adults (Arbuckle, Gold, & Andres, 1986; Christensen et al., 1996; Craik, Byrd, & Swanson, 1987; Erber & Szuchman, 1996; Hill, Wahlin, Winblad, & Bäckman, 1995; Hultsch, Hammer & Small, 1993; Luszcz, Bryan, & Kent, 1997; Mitchell et al., 2012; Stern, 2009; van Boxtel, Langerak, Houx, & Jolles, 1996). Specifically, these studies demonstrate strong associations between higher levels of education, later retirement age, more mentally demanding occupations, volunteering and greater performance on various cognitive tasks in healthy older adults (Fratiglioni et al., 2004; Karp et al., 2009; Lupton et al., 2010; Schaie, 1994; Shimamura, Berry, Mangels, Rusting, & Jurica, 1995; Yaffe et al., 2009). In addition there is evidence that higher levels of engagement in mentally stimulating leisure activities, such as crosswords, chess, puzzles, bridge, driving, flying, playing music or even visiting museums during midlife, are associated with self-reported decreased cognitive decline in older adulthood relative to individuals who self-report less engagement (Christensen et al., 1996; Kramer et al., 2009; Kramer & Willis, 2002; Mitchell et al., 2012; Park et al., 2007; Singh-Manoux, Richards, & Marmot, 2003; Scarmeas, Levy, Tang, Many, & Stern, 2001; Stern, 2009; Valenzuela & Sachdev, 2006a; Valenzuela & Sachdev, 2006b; Wilson et al., 2002b). Observational studies suggest that undertaking productive activities like gardening, cooking and knitting is also a good strategy for slowing the influence of cognitive decline in older adulthood (Verghese et al., 2006; Wilson et al., 2007).

**Longitudinal studies.** The level of evidence of the above-mentioned studies to make associations between the causes of cognitive ageing (or successful ageing) and levels of enrichment is limited. Longitudinal studies offer better insights because they admit fewer rival explanations of observed effects (or lack of effects) than cross-sectional evidence (Deary et al., 2009a; Scarmeas & Stern, 2003). Promisingly, longitudinal studies in older adults have also found that healthy older adults who participate in more intellectually challenging daily activities show less decline over time on various tests of cognitive performance (Hultsch et al., 1999; Mitchell et al., 2012; Small, Dixon, McArdle, & Grimm, 2012; Treiber et al., 2011; Yaffe et al., 2009).

***Longitudinal studies utilising more sophisticated statistics.*** Furthermore, there are a growing number of longitudinal studies implementing more sophisticated statistical techniques to consider rates of change across time. In addition, new techniques can explore predictors (*i.e.*, protective and risk factors) for rates of cognitive decline, including enriched lifestyles, as well as inter-individual differences in cognitive trajectories (Gold et al., 1995; Hultsch et al., 1999; Schooler & Mulatu, 2001; Treiber et al., 2011). The importance of considering changes in inter-individual rates of decline (or exploring the change in trajectories following cognitive training) in relation to protective and risk predictors – such as proxies for CR – is elaborated further below.

There is evidence of longitudinal cognitive performance gains from cognitive enrichment in older adults. Studies conducted by Hultsch and colleagues (1999) and more recently by Small and colleagues (2012) have all demonstrated cognitive performance gains. These researchers examined data from the Victoria Longitudinal Study (VLS), a study of middle-aged and older adults (*n* = 250, *n* = 952, respectively) using structural equation modelling (SEM) and latent change score models. They investigated whether lifestyle activities buffer normal ageing-related declines in cognitive performance across 6- and 12-year periods. Changes in intellectual activities were related to changes in cognitive functioning, consistent with the hypothesis that intellectually-engaging activities buffer individuals against decline on an array of cognitive variables. Decrements in lifestyle engagement were related to poorer cognitive functioning. Overall these findings suggest that the enriching lifetime experience through engagement in higher levels of activity is related to more successful ageing.

A recent statistically sophisticated study by Mitchell and colleagues (2012) adds further evidence for the positive impact of cognitive stimulating activities (*e.g.*, crossword puzzles) on longitudinal cognitive performance. It examines predictors of performance which, interestingly, highlights the necessity to continue to 'Use it' or 'lose it', to avoid or slow down the progression of ARCD. It emphasises change in cognitive performance is essential in order to demonstrate gains, as noted with plasticity theories highlighted earlier in the chapter. The authors looked at the effects of cognitive activity on cognitive trajectories over a period of 21 years using a number of outcome measures. They used a series of mixed effects models of data from four longitudinal studies of ageing: the Origins of Variance in the Oldest-Old: Octogenarian Twins Study (Octo-Twin; $n$ = 572 at baseline); the Long Beach Longitudinal Study (LBLS; $n$ = 561 at baseline); the Seattle Longitudinal Study (SLS; $n$ = 1649 at baseline), as well as the VLS ($n$ = 1011 at baseline). The *predictors* of cognitive outcomes examined, included an investigation of change in cognitive activity from baseline. Results indicated that change in activity was associated with relative change in cognitive performance trajectories in specific cognitive domains. In particular, increases in cognitive activity from baseline were associated with better than expected cognitive performance and, conversely, activity decrease was associated with worse than expected performance. When baseline cognitive activity level was used as a predictor of cognitive trajectories, it did not predict cognitive decline over time. Thus the findings suggest that, individuals who increased cognitive activity may thereby effectively reduce their level of ARCD. This finding is consistent with the extant positive plasticity literature, Salthouse's disuse perspective and indeed the concept of older adults' needing to 'Use it' or 'lose it'.

***Longitudinal studies predicting inter-individual differences in performance trajectories.*** Whilst the above studies offer good insights into the exploration of cognitive trajectories across time, there are few studies adequately exploring inter-individual differences in cognitive trajectories across time, and their predictors [9] (Jones et al., 2005; Terrera et al., 2010; Yaffe et al., 2009).

Terrera and colleagues (2010) described predictors of inter-individual differences in longitudinal trajectories of cognition in older adults ($n$ = 2053) using growth mixture models (GMM; Muthén, 2002). They fitted data from baseline, two-, seven- and nine-year follow-up interviews of individuals participating in the Cambridge City Over 75 Cohort study (CC75C). They identified heterogeneity in cognitive change, specifically demonstrating three cognitive trajectories or classes, basing performance on the Modified Mini-Mental State Examination (mMMSE): a slow decline (41% of the sample; those classified as being within the relatively successful ageing group); an accelerating decline from a baseline of cognitive impairment (54%); and a steep

[9] Further discussion of predictors and consideration of cognitive trajectories across time to better elucidate cognitive changes is discussed below from theoretical perspectives (*i.e.*, the differential preservation versus preserved differentiation hypotheses; Salthouse, 2006; Salthouse, Babcock, Skovronek, Mitchell, & Palmon, 1990).

constant decline also from a baseline of cognitive impairment (5%). Predictors of performance included education to explore different influences within each class. The authors demonstrated the protective effect of education was strong in those exhibiting successful ageing, where the rate of cognitive decline was slower in those with higher education levels (*i.e.*, their cognitive trajectories had a smaller gradient). The study also demonstrated the effect of sex on cognitive performance of those exhibiting successful ageing with females demonstrating greater decline.

A similar investigation by Yaffe and colleagues (2009) over a period of eight years, explored predictors of ARCD in 2,509 well-functioning older men and women, as noted in Chapter 2. They also used random effects models of the mMMSE performance to demonstrate heterogeneity of the manifestation of cognitive decline but also investigated predictors of this decline. Three groups were identified: major decliners (16% of the sample); minor decliners (53%); and maintainers who exhibited no decline on global cognitive function (30%). The 'no decline' group results were considered to demonstrate older adults who were exhibiting successful ageing. This successful ageing group had a unique profile differentiating them from the other groups, with factors useful in predicting cognitive performance including education and literacy levels. Individuals with a high school education or greater, as well as greater literacy levels, were more likely to be a member of the maintenance subgroup than of the major cognitive decline subgroup.

Finally, Jones and colleagues (2005) modelled recall and learning on the Rey Auditory Verbal Learning Test (RAVLT) for individuals in the largest cognitive training study to date, the Advanced Cognitive Training for Independent and Vital Elderly (ACTIVE; Ball et al., 2002; Rebok et al., 2014; Willis et al., 2006) study using latent growth curve techniques. They found that individual differences in learning were related to greater verbal knowledge using the multiple choice vocabulary test from the Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, Harman, & Derman, 1976), in addition to older age and less education being associated with lower memory performances.

Together these longitudinal studies, particularly those utilising sophisticated statistics demonstrate that measures of cognitive enrichment, as well as proxies for CR (including education and indices of intelligence), positively impact on inter-individual cognitive trajectories across time. That is, an intellectually-engaged lifestyle overall has been shown to act as a buffer against cognitive decline and may be useful in contributing to more successful ageing. This seems to add further weight to the ‘Use it or lose it’ and associated theories.[10]

## Limitations of ‘Use It or Lose It’ Hypothesis and Associated Theories

### Current Lines of Investigation

Whilst there is substantial evidence of a link between cognitive enrichment and better cognitive performance, these studies are not without limitations. These limitations are not often considered by the media and popular culture, which has embraced the concept, leading to an infultration of the term ‘Use it or lose it’ into vernacular (Daffner et al., 2010).

There is a distinct lack of quality studies investigating levels of mental ‘exercise’ (Salthouse, 2006). For example, the statistical power of epidemiological, correlational and cross-sectional studies is not strong, as previously noted. The National Health &

[10] Utilisation of similar longitudinal statistical methods has also been demonstrated as useful in exploring other predictors of successful ageing, including cognitive training, and is discussed in more detail in Chapter 4. Further elaboration on similar statistical techniques for analysing cognitive trajectories is included in Chapter 5.

Medical Research Council (NHMRC) rates epidemiological and related studies as providing a low level of evidence (level IV) whereby no causal conclusions can be drawn (*e.g.*, Lupton et al., 2010). Cross-sectional correlations are not causal, and may be brought about by a third, confounding, variable or set of variables.

Similarly, it is difficult to separate out the effect of selection, selective attrition and causal directionality, whereby individuals who may already have greater cognitive abilities (either inherent or baseline performances) are more cognitively active, seek out stimulation and thus exhibit less ARCD (Depp et al., 2012; Deary et al., 2009a; Hertzog et al., 2009; Salthouse, 2009; Scarmeas & Stern, 2003; Willis et al., 2009).

A number of methodological issues still remain concerning longitudinal studies (Deary et al., 2009a; Scarmeas & Stern, 2003; Salthouse, 2006). As noted above, there is the need to consider cognitive trajectories across time to better elucidate cognitive changes. Salthouse and colleagues (Salthouse, 2006; Salthouse et al., 1990) highlight moving away from considering cognitive performance at a single time point and instead considering *rates* of change when investigating ‘Use it or lose it’ and associated theories. As noted, few studies take this approach statistically. Considering rates of change would lead to a better understanding of factors contributing to inter-individual differences in the expression of ARCD. Salthouse and colleagues highlight and describe the importance of these considerations within the ‘differential preservation’ and ‘preserved differentiation’ hypotheses (Salthouse, 2006; Salthouse et al., 1990).

**Differential Preservation Versus Preserved Differentiation**

The *differential preservation* hypothesis, also referred to the ‘*differentiation preservation*’ hypothesis (Salthouse, 2006; Salthouse et al., 1990; Tucker-Drob &

Salthouse, 2011), describes individual differences in cognitive ageing trajectories resulting from different levels of a particular factor (or of multiple factors), such as mental exercise. These factors *interact* with the cognitive ageing process and produce greater *between-person* variability in cognitive performance, as is seen with increasing age (Salthouse, 2012). Importantly, this hypothesis posits that these factors predict the *rate* of decline. Specifically, the differentiation preservation hypothesis emphasises the impact of predictive factors on the *preservation* of cognitive function. This means that preservation of performance is the *differential* although, of course, an increase or decrease in cognitive function are alternatives.

The left panel of Figure 1 provides an example of differential preservation. In this example consider mental activity as a potential protective factor. An individual with a higher level of mental activity would show a more successful ageing trajectory. That is, their performance trajectory would have a flatter gradient and the individual would demonstrate less cognitive decline; in other words, they would be showing a *preservation* of their cognitive function. An individual at an average level of a mental exercise would show a moderate declining trajectory, whilst an individual with a low level of mental activity would demonstrate the greatest negative gradient of cognitive change.[11]

In fact the individual with the lowest level of a protective factor may pass a threshold of significant functional impairment. Thus, when considering trajectories of normal ARCD, an individual's low levels of a protective factor would impact the trajectory of

[11] A fourth alternative could be that a predictor would influence the rate of ageing to an extent where cognitive performance had a positive trajectory. This would represent the most exciting outcome in relation to factor's impacting on ARCD. This is not depicted in the figure.

decline to bring them to the critical significant functional impairment threshold sooner than those with higher levels of the protective factor.

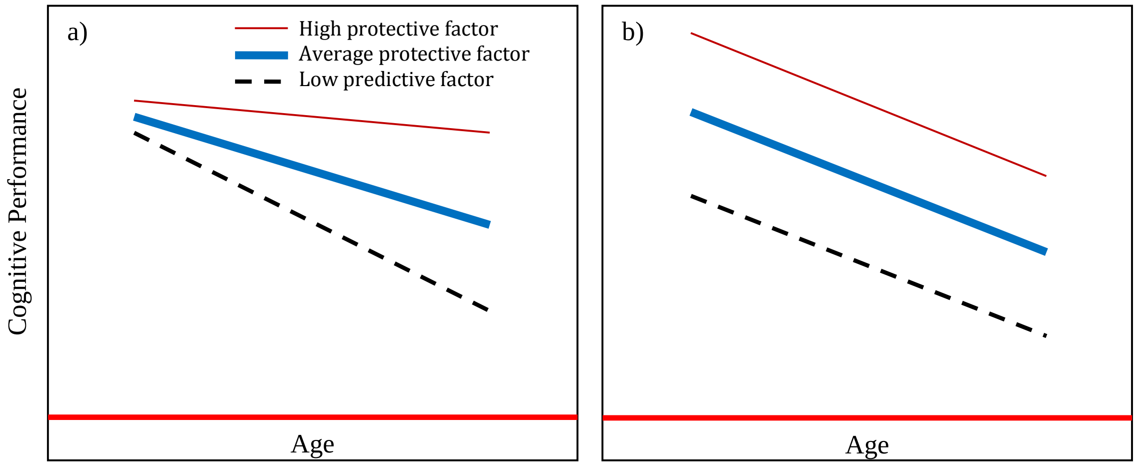


*Figure 1.* a) Differential preservation and b) preserved differentiation hypotheses. This illustration demonstrates the different trajectories of hypothetical individuals as described by both theories. The horizontal line depicts a diagnostic threshold beyond which daily functioning is significantly compromised.

Alternatively the *preserved differentiation* hypothesis (Salthouse, 2006; Salthouse et al., 1990; Tucker-Drob & Salthouse, 2011), suggests that whilst individuals may differ in their level of a predictive factor, they *do not* differ with respect to the *rate* of decline. That is, the inter-individual differentiation is preserved, or fixed, across time (Tucker-Drob & Salthouse, 2011). Heterogeneity of cognitive functioning with age is also demonstrable due to pre-existing differences in a protective factor. In other words, the slope of cognitive decline is similar across individuals; however, differences in performance at a *specific time point* are present because individuals begin their decline at different levels of cognitive ability (Salthouse, 2006; Salthouse et al., 1990). Thus following this hypothesis, individuals who go on to reach the functional impairment threshold earlier do so because they have begun adulthood closer to the critical functional threshold.

To clarify these concepts in relation to mental activity and the ‘Use it or lose it’ hypothesis, the key difference between the differentiation preservation and preserved differentiation hypotheses is that the former considers mental activity as being protective against ARCD, whilst the latter views an individual’s current level of cognitive performance as being at least partly due to a manifestation of their pre-existing level of mental ability (*e.g.*, intelligence). The differential preservation hypothesis would suggest that baseline characteristics and/or cognitive training provide enrichment to build the neurobiological mechanisms to counter ARCD. The preserved differentiation hypothesis would suggest that a minimum level of mental strength is needed to better resist ARCD.

In relation to the preserved differentiation hypothesis, Salthouse (2006) and others (*e.g.*, Park et al., 2007; Hertzog et al., 2009) add that whilst the rate of change may not vary across individuals given differences in levels of predictive factors, this does not discount that purposeful manipulation of a higher level of a protective factor (*e.g.*, cognitive training) may be implemented. Such an action could broadly and positively impact the absolute level of performance at that point in time. Theoretically this improved performance could then be maintained over time, thereby altering the trajectory of ageing from the start of an intervention, such as cognitive training (Park et al., 2007). Such an effect may be of practical significance to keep an individual from passing the critical threshold of functional impairment. Indeed, theoretically, trajectories of change can manifest rather than just change as an absolute level of performance, immediately after the intervention; thus there could still be an interaction between normal ageing trajectories and cognitive training.

In sum, the differential preservation and preserved differentiation hypotheses highlight that it is predictors of inter-individual differences in cognitive *trajectories* in older adults that are the crucial consideration, rather than an assessment of cognitive function at a specific time point. This is useful when considering those exhibiting symptoms of ARCD. Importantly both theories provide support for the concept that *environmental enrichment* (EE) can positively alter these trajectories. It is, therefore, also possible that experimental manipulation can alter these trajectories. The theories offer a clear basis from which inter-individual differences in training outcomes can be viewed, in relation to the impact (or lack thereof) of predictors on cognitive performance trajectories following training. This concept of the rate of cognitive trajectory changes is further discussed in relation to the impact of cognitive training in the following chapter.

**Limitations to Concept of Cognitive Enrichment in Research**

There are a number of other issues when considering cognitive enrichment literature. Firstly, measures of cognitive enrichment lack an operationalised definition. A diverse array of activities has been argued to be mentally stimulating (*i.e.*, they involve a cognitive process or cognition to some extent), and these activities have been used in several different ways to scale potential cognitive enrichment (Jopp & Hertzog, 2007). However what exactly constitutes mental exercise is not clear or always consistent. Watching television is a prime example; it has been included as a positive behaviour and on several general scales of activity (Crowe, Andel, Pedersen, Johansson, & Gatz, 2003; Wilson et al., 2002b. Cross-sectionally watching television has also been identified as a behaviour negatively associated with cognition (Hertzog et al., 2009; Jopp & Hertzog, 2007), and has been associated longitudinally with an increased risk of developing cognitive impairment (Wang et al., 2006).

Measures of cognitive enrichment are also lacking as they largely rely on retrospective self-reports. It is unclear how best to quantify engagement with the task and how to account for task difficulty (Depp et al., 2012; Frick & Benoit, 2010). For example, social stimulation is also considered to be an important constituent of cognitive enrichment, but is difficult to measure. Furthermore, levels of social stimulation might vary, not only in amount but also in quality (Hertzog et al., 2009).

Secondly, cognitive abilities (such as proxies for CR education and IQ) may influence lifestyle choices in adulthood and the exhibition of individual differences in cognitive decline. Described as a 'multiplier effect', an individual with higher IQ may seek out a stimulating environment. For example, higher levels of education may lead to a more challenging occupation and also to a greater engagement in exciting leisure activities, which are both more cognitively stimulating (Kramer et al., 2004). A 'multiplier effect' could create significant cognitive gains even with small environmental changes (Willis et al., 2009). With regard to health variables, intelligence measured at childhood is associated with health variables and health risk factors in middle and older age (Deary et al., 2009b). The onset of chronic disease (*e.g.*, hypertension, cardiovascular disease) also has a major impact on the maintenance of fluid intelligences in later life (Schaie, 2008). In addition, cognitive EE leads to concomitant health effects, lower levels of depression, stress reduction and improved vascular health, which independently may potentially affect the expression of ARCD (Fratiglioni et al., 2004; Lövdén et al., 2010).

Cognitive reserve is also thought to influence an individual's receptiveness to intervention (Baltes, 1987; Boron et al., 2007a; Katzman, 1993). While controversial, this thesis supports the viewpoint that proxies for CR can measure an individual's overall learning potential and capacity to implement training protocols. The concept has been highlighted with regard to the possible *synergistic* effects of other cognitively stimulating environmental experiences, with education and/or intelligence (*e.g.*, Stern, 2009). Different prior levels of CR may thus interplay with cognitive interventions and therefore alter the effect of cognitive outcomes. That is, individuals with greater levels of CR may profit to a greater degree from further active attempts to achieve greater cognitive performance. This may be because they have a greater capacity to learn and implement training protocols (Bagwell & West, 2008; Garlick, 2002; Lövdén et al., 2010; Hill et al., 1995; West & Hastings, 2011). This is further discussed with regard to predictors of cognitive training outcomes in Chapter 4.

Finally, there may also be synergistic effects of stimulating activities and the constituents of these activities (Agrigoroaei & Lachman, 2011; Stern, 2009). That is, a combination of stimulating experiences through a number of avenues (*e.g.*, social and cognitive stimulation) and/or from a number of contexts could result in additive cognitive benefits (or even effects greater than the sum of their parts).[12] A study by Karp and colleagues (2006) investigated leisure activities with multiple constituents, including physical, mental and social activity and the activities' relationship with the onset of dementia. In the study of 800 participants aged 75 and older, those who were more active physically or mentally, or more socially engaged, had a lower risk of developing dementia. Those whose activities combined two or three of the components had more of a reduced risk. Thus, there are additive effects of multiple beneficial components of stimulating leisure activities.

On balance, however, the available evidence favours the hypothesis that maintaining an intellectually-engaged lifestyle through a variety of avenues leads to higher

[12] Indeed, this is what proponents of *multidomain training paradigms* suggest, discussed in Chapter 4.

cognitive activity, which in turn promotes more successful cognitive ageing (Hertzog et al., 2009).

### Summary

In sum, there are three concepts in the literature that are heuristically useful in the exploration of individuals who more successfully age versus those exhibiting less resistance to ARCD. These include CR, and plasticity theories, and the 'Use it or lose it' hypothesis.

The term CR was introduced in the neuropsychological literature as a way to explain inter-individual differences between older adults' cognitive performance in the presence of pathology, which can be extended to ageing. The neurobiological concept of plasticity can be used to explain the mechanisms through which neural and cognitive changes can occur (*i.e.*, neuroplasticity and cognitive plasticity).

Both theories emphasise two key factors relating to inter-individual differences in the expression of ARCD. First, that individuals may possess an adaptive inherent capacity, derived through biological mechanisms. Secondly that environmental experience, especially the enriching effects of positive environmental experiences can increase cognitive performance.

Finally, the central overarching and predominating hypothesis, popularly known as 'Use it or lose it' was introduced. It also emphasises key features of the three concepts. That is, the need for cognitively enriching environments and lifestyles, in which there is complexity and novel stimulation. Without such stimulation, cognitive function is 'lost', just as Salthouse's disuse perspective (1991) purports.

Whilst not without their deficits, a number of studies, including imaging, epidemiological, cross-sectional, correlational and longitudinal are all generally supportive of the overarching thrust of the ‘Use it or lose it’ hypothesis, CR and plasticity theories. The need to investigate longitudinal rates of change was highlighted through discussion of the differential preservation and preserved differentiation hypotheses. These two hypotheses highlight the need to consider predictors of inter-individual differences and rates of change, in older adults. This is useful when considering those exhibiting symptoms of ARCD. Importantly both theories provide support for the concept that inherent ability or enrichment can positively alter these trajectories. Also noted was the lack of an operational definition of cognitive enrichment, and the possibility of a complex synergistic interplay of factors in relation to cognitive enrichment itself, and inherent ability, such as education and intelligence.

Nonetheless, overall the chapter highlights that cognitive functioning is not necessarily restricted to an inherent predisposition, and that cognitive performance may instead be at least partially promoted by environmental factors.

## LITERATURE REVIEW

# Chapter 4

# Cognitive Training

The emphasis on enriching environments with cognitive reserve (CR) and plasticity theories, and the 'Use it or lose it' hypothesis to bolster cognitive performance in older adults has led to an explosion in cognitive training paradigms. That is, studies in which there is direct manipulation of the environment via application of cognitive intervention.

This chapter discusses the mixed evidence of the efficacy of cognitive training as a paradigm, which includes some limitations such as a lack of demonstrated generalised effects of training to cognitive domains less proximal to trained skills and a paucity of studies with long-term follow-up. Importantly, it also notes the limited research investigating inter-individual differences in longitudinal trajectories of change following cognitive training. Ignoring this heterogeneity obscures the accuracy of training results. The chapter will also show that identifying predictors of cognitive *trajectories* following training is required to adequately examine the 'Use it or lose it' hypothesis and associated theories.

### Defining Cognitive Training

*Cognitive training* falls under an overarching concept of cognitive remediation; the application of an intervention designed to mediate deterioration in cognition**.** In the ageing context, cognitive remediation is an intervention method designed to reduce, or reverse ARCD. Cognitive training involves the structured instruction of *strategies* and guided practice on various tasks to train cognitive functions, for example, fluid processing such as memory, attention, speed of processing and executive functions (Clare, Woods, Moniz-Cook, Orrell, & Spector, 2003).

Also falling into the realm of cognitive remediation, but are generally considered distinct from cognitive training, is cognitive stimulation and cognitive rehabilitation. *Cognitive stimulation* refers to the use of a wide range of activities, which *non-specifically* enhance cognitive and social functioning (Buschert, Bokde, & Hampel, 2010; Tesky, Thiel, Banzer, & Pantel, 2011). *Cognitive rehabilitation* is generally considered in the context of disease or injury, *e.g.*, MCI, dementias, stroke or acquired brain injury (Acevedo & Loewenstein, 2007; Belleville, 2008; Clare & Woods, 2003; McLellan, 1991; Medalia & Richardson, 2005; Sitzer et al., 2006).[13] Cognitive rehabilitation is also often considered to more specifically address particular cognitive deficits. In contrast, cognitive training can be considered to cover both specific and broad domains of cognitive function (Papp et al., 2009). The current review and empirical study focuses on cognitive training interventions.

### Subtypes of Cognitive Training

Cognitive training has many subtypes, including strategy- and process-based training and multi-domain training. These cognitive training types can be individual or group-based. Computer programs can also be used.[14] Cognitive training interventions are thought to use plasticity and flexibility mechanisms to improve cognitive performance (Brehmer, Li, Muller, von Oestzen & Lindenberger, 2007; Hertzog et al., 2009;

[13] As noted in chapter 2, the present study's consideration of the cognitive ageing process as being 'normal' and not a representation of a pathological process remains a contentious issue (*cf.* Wilson et al., 2010).

[14] Given the scope of the current review, only group-based interventions will be discussed, particularly because interactive group programs have also been shown to increase efficacy of training programs (Verhaeghen, Marcoen, & Goossens, 1992).

Lövdén et al., 2012; Lustig & Flegal, 2008; Lustig et al., 2009). They align with 'Use it or lose it' to address cognitive and functional challenges seen with ageing (Buitenweg et al., 2012; Lustig et al., 2009; Park & Reuter-Lorenz, 2009; Reuter-Lorenz & Park, 2010; Stern, 2009). Cognitive training subtypes are further elucidated below.

**Strategy-based training.** Strategy-based training approaches teach and facilitate practice of one or several specific techniques for performing a particular cognitive task. This is usually with the explicit goal to maintain or enhance strengths and/or to adapt to or increase performance in a specific cognitive weakness (Lustig et al., 2009; Park et al., 2007; Salthouse, 1991; Willis & Schaie, 2009). Strategy-based training programs often target memory and centre on *mnemonics*. Mnemonics (internal strategies) are cognitive techniques involving the organisation of items into meaningful groups to assist with encoding, retention and learning of new information (Brooks, Friedman & Yesavage, 1999b). Techniques used include imagery, name and face associations, as well as the method of loci, which involves visualising items in a sequence of specific, well-learned locations (Fairchild et al., 2012; Gatz, 2005). Mnemonics are considered 'top down' approaches, involving higher order cognitive processing, such as EF. They are thought to increase prefrontal white matter and caudate nucleus activation, which are particularly vulnerable to ageing processes (Park et al., 2007; Raz, 2009).

In addition to mnemonics, external strategies can also be taught in strategy-based cognitive training. External strategies involve the use of practical aids to compensate for weaker cognitive processes, *e.g.*, writing notes, using calendars (Acevedo & Loewenstein, 2007; Belleville, 2008; Clare & Woods, 2003; Medalia & Richardson, 2005; Sitzer et al., 2006). Instruction is provided to apply them in day-to-day contexts.

**Process-based training.** Process-based cognitive training (also referred to as behavioural, practice-based or 'bottom up' training) is a newer generation of training than strategy-based interventions. Process-based training involves participants engaging in mental activity resulting from complex and novel tasks that also often involve repetition and structured experience in training situations (Mohs et al., 1998; McDougall, 1999; Verhaeghen et al., 1992). They involve no explicit strategy training.

Process-based training is thought to be more likely than strategy-based training to use inherent cognitive plasticity mechanisms to contribute to CR (Gatz, 2005; Hertzog et al., 2008a; Jaeggi, Buschkuehl, Jonides & Perrig, 2008; Lövdén et al., 2010; Lustig et al., 2009; Stern, 2002). That is, performance gains result from the development of different strategies, response mappings and perceptual expertise, rather than use of specifically taught compensatory strategies (Hertzog et al., 2009; Lustig et al., 2009). Process-based training therefore has a greater likelihood of transfer of benefits to cognitive domains not specifically targeted (Park et al., 2007).[15]

Like strategy-based paradigms, training is often targeted to a specific cognitive domain, for example, memory, EF or a combination of cognitive processes (Basak, Boot, Voss, & Kramer, 2008; Bherer et al., 2005; Karbach & Kray, 2009; Kramer, Larish, & Strayer, 1995; Lustig et al., 2009; Mahncke et al., 2006b; Zelinski, 2009). Memory training can involve retrieval and encoding practice (Bissig & Lustig, 2007; Jennings & Jacoby, 2003), training working memory, *e.g.*, *n*-back tasks (Buschkuehl et al., 2008; Dahlin, Stigsdotter Neely, Larsson, Bäckman, & Nyberg, 2008; Jaeggi et al., 2008; Li, 2003; Olesen et al., 2004), attentional control (*e.g.*, divided attention and task-switching paradigms; Hertzog et al., 2008a; Noack et al., 2009).

[15] Specificity and generalisability of training efficacy will be further explored later in this chapter.

**Multi-domain training.** Multi-domain programs, also known as multi-modal or multi-factorial programs, include multiple components, incorporating both strategy- and process-based approaches (Hertzog et al., 2008; Noack et al., 2009). These programs offer a holistic approach, with novel tasks to optimise program efficacy and increasing the potential for transfer effects compared with strategy- or process-based interventions alone (Cheng et al., 2012; Park et al., 2007). Multi-domain training paradigms are also considered more likely to produce longer-term benefits (Buitenweg et al., 2012).[16]

Multi-domain approaches emphasise continuing intellectual, social and physical enrichment.[17] They target factors seen to contribute to successful ageing (APA, 2000; *cf.* definitions by Butler, 1991; Rowe & Kahn, 1987; WHO, 2002). Multidomain programs can include an emphasis on mnemonics and provide psychoeducation (*e.g.*, regarding memory changes in normal ageing, and the influence of lifestyle factors, such as diet, sufficient sleep, stress and mood management, and the general promotion of cardiovascular health; Floyd & Scogin, 1997; Gatz, 2005; Hertzog et al., 2008; Hess, 2005; Hohaus, 2007; Hultsch et al., 1999; LaRue, 2010; Schneider & Yvon, 2012; West, Bagwell, & Dark-Freudeman, 2008; West & Hastings, 2011; Woolverton, Scogin, Shackleford & Duke, 2001). Group-based multi-domain training offers social stimulation. Social engagement has been shown to have a positive effect on cognitive functioning maintenance (*e.g.*, Barnes, Mendes de Leon, Wilson, Bienias, & Evans, 2004), whilst social isolation leads to declines in cognition and functional capacity (Green, Rebok, & Lyketsos, 2008b). Relaxation techniques and homework tasks can

[16] Long-term retention of training cognitive gains is further explored later in this chapter.

[17] As previously noted, the current review will not discuss physical enrichment. There is, however, an expansive array of work in both human and animal populations on the contributions of aerobic exercise on cognition in ageing, either alone or in combination with the cognitive training methods outlined above (*cf.* Oswald, Gunzelmann, Rupprecht, & Hagen, 2006; Lustig et al., 2009).

also be incorporated (Hohaus, 2007; Stigsdotter-Neely & Bäckman, 1993; Stigsdotter & Bäckman, 1989; Stigsdotter-Neely & Bäckman, 1995). The efficacy of multi-domain approaches has been demonstrated in the literature, albeit with some having only a small number of participants (*e.g.*, Cheng et al., 2012; Hohaus, 2007; Hultsch et al., 1999; Jones et al., 2006; Lustig et al., 2009; Martin et al., 2011; Stigsdotter-Neely & Bäckman, 1993; Park et al., 2007; Stigsdotter-Neely & Bäckman, 1989; Stigsdotter-Neely & Bäckman 1995).

## Efficacy of Cognitive Training Paradigms

Whilst there is some evidence to support the efficacy of multi-domain training in the cognitive training literature, there is a lack of clarity as to the extent and nature of effectiveness of cognitive training interventions overall on cognitive performance in older adults. This is due to a number of issues with current research as noted above. Table 1 below summarises the evidence (or lack thereof) of specificity, generalisability and longitudinal effects of cognitive training.

Table 1.
*Specificity Versus Generalisabilty and Evidence of Long-term of Training Effects*

| Authors | Study type | *n* | Evidence of specificity and generalisability of training | Evidence of long-term effects |
|---|---|---|---|---|
| Verhaeghen, Marcoen and Goosens (1992) | Meta-analysis (49 mnemonic training studies) | 1,539 | Moderate to large specific pre-post training gain for memory training (Cohen's $d$ = 0.73) Small specific training gains *vs.* placebo and control groups (Cohen's $d$ = 0.37 and 0.38, respectively). No generalised effect measured | n/a |
| Valenzuela and Sachdev (2009) | Meta-analysis (7 RCTs) | 3,194 | Large specific training *vs.* wait-list control (weighted mean difference = 1.07). Generalised effects were evidenced in some studies. | "Strong" evidence (p.179): Average effect size from RCTs with ≥2-years follow up within the 95% confidence interval of those with <2-year follow up (weighted mean difference =1.02 versus 1.16, respectively).[18] |
| Papp, Walsh and Snyder (2009) | Meta-analysis (10 RCTs) | 3,941 | Small specific effects pre-post across most measures (Cohen's $d$ = 0.16; range 0.138–0.186). Generalised effects were only reported in two of the seven studies. | Insufficient evidence |

[18] The authors noted that quality of studies considered to be low.

| Authors | Study type | *n* | Evidence of specificity and generalisability of training | Evidence of long-term effects |
|---|---|---|---|---|
| ACTIVE study (Ball et al., 2002; Rebok et al., 2014; Willis et al., 2006) | Largest RCT to date (single blind) | 2,832 | Small specific pre-post effect on memory performance compared to controls (effect size = 0.257).<br>Small to medium specific post-test effect on reasoning performance compared to controls (effect size = 0.480).<br>Small specific increase in speed of processing compared to controls (*i.e.*, decreased performance effect size = −1.463).[19]<br>Generalised effects were not demonstrated. | Long-term specific training effects at one-, two-, five- and 10-year follow-ups:<br>Small effect on memory: one and two years post training only (effect sizes 0.212 and 0.174, respectively).*<br>Medium to small specific long-term effect across time on reasoning training (effect sizes 0.402, 0.257, 0.26 and 0.23, respectively).<br>Large to moderate increased speed of processing performance level (effect size 1.212, 0.867, 0.76 and 0.66, respectively). |
| Martin and colleagues (2011) | Meta-analysis (36 RCTs of cognitive interventions) | 2,229 | Specific training effects for memory *vs.* no-contact controls.<br>“not sufficient information” to analyse generalised effects (p. 48). | Insufficient evidence |

*Note*: RCT = randomised control trial.
* Authors noted decline in performance between baseline to year 5, although this decline was smaller than the control group. At 10 years, however, memory training effects were no longer maintained (Rebok et al., 2014).

[19] Effect of training was defined as (trained mean − control mean at later time) − (trained mean − control mean at baseline) (Ball et al., 2002).

Table 1 shows that largely studies suggest cognitive training may produce specific cognitive functioning gains in older people with group data (Ball et al., 2002; Martin et al., 2011; Rebok et al., 2014; Willis et al., 2006; Verhaeghen et al., 1992). That is, training effects are largely only seen in cognitive domains more proximal to trained skills. For example, in the ACTIVE study (Ball et al., 2002; Rebok et al., 2014; Willis et al., 2006), memory training only produced memory improvements. The memory trained group demonstrated no gains in performance in reasoning or speed of processing (in fact, the authors reported that they showed decline). Also referred to as 'near transfer' effects, specificity of effects is supported by others (*e.g.*, Brehmer et al., 2008; Noack et al., 2009; Park et al., 2007; West & Hastings, 2011; Zelinski et al., 2011). Effect sizes (e.g. Cohen's *d* and weighted mean differences) across the studies vary, and are relatively small however, and to date there has also been a strong focus on memory outcomes only (Martin et al., 2011; Verhaeghen et al., 1992).

In contrast, there is a lack of demonstrated *generalised* effects (or 'far transfer') of training (*i.e.*, transfer of training benefits to cognitive domains less proximal to the trained skill). In cases where there is transfer, these are often small (Ball et al., 2002; Hertzog et al., 2009; Jones et al., 2006; Lustig et al., 2009; Papp et al., 2009; Stigsdotter-Neely & Bäckman, 1993; Tucker-Drob & Salthouse, 2011; Willis et al., 2006). Generalisation of cognitive gains across multiple cognitive domains is seen as an essential outcome of such programs indicating wider success of training (Gatz, 2005). Martin and colleagues (2011) showed that there was only a sufficient number of studies in the memory domain, not other cognitive domains, such as executive functioning. Executive functions are well known to decline in population studies of ARCD. Thus there is now a need for well-designed studies to examine generalised effects, including those beyond memory, into domains also noted to decline with ageing.

Finally, there is a paucity of studies with long-term follow-up. Long-term maintenance (*i.e.*, durability) of induced improvements following the discontinuation of training should be sought when assessing the efficacy of a program (Deary et al., 2009a; Gatz, 2005; Gross et al., 2011; Lustig et al., 2009; Raz, 2009). Of those studies that do assess longitudinal outcomes, a number only demonstrate immediate performance gains, and poor continuance of cognitive gains over time (Gatz, 2005; Rebok et al., 2007; Valenzuela & Sachdev, 2009; Verhaeghen et al., 1992). Thus, there is insufficient evidence of long-term benefits from training (Papp et al., 2009; Martin et al., 2011).

Overall, it can be seen that clear answers regarding evidence for the specificity, generalisability and long-term efficacy of cognitive training remains elusive. Further quality investigations into generalised and longitudinal training effects are warranted. These issues create ambiguity as to whether the evidence supports the ‘Use it or lose it’ hypothesis and associated theories. That is, whether through the use of cognitive resources via training intervention cognitive decline can be decreased or even reversed.

## Issues with Conventional Approaches to Assessing Training Effects

### Inter-individual Differences in Cognitive Training Outcomes and Trajectories of Change

A further limiting factor of previous research into cognitive training effects is inadequate consideration of inter-individual differences in cognitive training outcomes, and a lack of consideration of *rates* of cognitive change (Buitenweg et al., 2012; Jolles & Crone, 2012; Karbach & Schubert, 2013; Leoutsakos et al., 2012; Lövdén et al., 2012; Martin et al., 2011; Park, 2007; Terrera et al., 2010; West & Hastings, 2011).

There is growing evidence of inter-individual differences in training responsiveness. Training effects seen in a specific individual may be substantially different from group effects (Martin et al., 2011). Inter-individual differences can be quite large (Bissig & Lustig, 2007; Yesavage et al., 1988). Ignoring heterogeneity in training responses may obscure the accuracy of training study results and raise questions as to the validity of training results (Ball et al., 2002; Baltes & Kliegl, 1992; Boron et al., 2007a; Duncan et al., 2002; Fairchild et al., 2013; Fandakova et al., 2012; Langbaum et al., 2009; Rosen & Yesavage, 2012; Schaie, Willis, Hertzog, & Schulenberg, 1987; Willis & Nesselroade, 1990; Willis et al., 2006; Zelinksi et al., 2007).

The ACTIVE study (Ball et al., 2002; Rebok et al., 2014; Willis et al., 2006) provides evidence of the heterogeneity of training responsiveness in older adults. Although the ACTIVE study demonstrated overall significant training effects for the experimental group, training effects were only found in a small proportion of individuals trained. Specifically, only 26% of participants' in the memory training group demonstrated significant improvement in subsequent memory testing (Ball et al. 2002). Eighty-seven percent of speed-trained and 74% of reasoning-trained individuals showed improvement immediately following intervention; that is, not all of those trained profited from the interventions.

The existence of inter-individual differences highlights issues with the use of conventional group-level statistics and the need for more appropriate alternative statistical techniques. Most studies employ group-level, variable-centred statistics, for example, Analysis of Variance (ANOVA) and multiple regression (Connell & Frye, 2006; Langbaum et al., 2009; Willis & Schaie, 1987). These statistics examine average outcome variable scores and the predictive relationships between independent and dependent variables, *e.g.*, time versus cognitive performance (Connell & Frye, 2006; Jung & Wickrama, 2008). Whilst inter-individual differences are considered in these traditional approaches, the differences are treated as error variance. This error variance may, however, contain valuable information about change. Conventional statistics therefore may obscure information regarding the heterogeneity of individual performances within a sample exhibiting robust change (Connell & Frye, 2006; Langbaum et al., 2009; Willis & Schaie, 1987).

Conventional studies' statistical approaches are also not ideal in the training context because, ideally, the *process* of change needs to be examined. That is, examination of growth trajectories across time is warranted (Baltes & Nesselroade, 1979; Baltes & Kliegl, 1992; Collins & Horn, 1991; Kliegl et al., 1989; Langbaum et al., 2009; Martin et al., 2011; Muthén, 2004; Salthouse, 2006; Terrera et al., 2010). The effectiveness of a training intervention needs to be assessed by the extent to which it is capable of altering the normative growth *trajectory* (Baltes & Kliegl, 1992; Kliegl et al., 1989; Langbaum et al., 2009; Muthén, 2004; Salthouse, 2006). Conventional approaches only consider change at a specific time point, *e.g.*, 'responders' versus 'non-responders' (Deary et al., 2009a; Hedden & Gabrieli, 2004; Nagin & Odgers, 2010; Terrera et al., 2010). Without consideration of rates of change these more commonly-used statistics may lead to ambiguity in identification of treatment effects.

Clinically, identifying distinct trajectories provides information about participants' cognitive changes across time (*e.g.*, when changes are likely to occur, plateaus in performance *etc.*). Such information would also allow for more cost effective individual selection for training programs to maximise the cognitive gains achievable longer-term. Thus, overall inadequate experimental methodology, including limitations of conventional statistical approaches have contributed to the lack of clarity within the training literature regarding the efficacy of training programs at the group level. A move away from conventional investigative paradigms is required (Hedden & Gabrieli, 2004; Martin et al., 2011; Terrera et al. 2010).

Group-based growth modelling addresses the above-mentioned limitations of conventional statistical approaches and experimental paradigms. GBGM provides essential, more valid information about inter-individual differences in longitudinal growth trajectories (Deary et al., 2009a; Hedden & Gabrieli, 2004; Lindenberger & von Oertzen, 2006; Lövdén et al., 2010; Nagin & Odgers, 2010; Nelson & Dannefer, 1992; Papp et al., 2009; Park et al., 2007; Raz, 2009; Schaie, 1994; Terrera et al., 2010; Valenzuela & Sachdev, 2009; Yaffe, 2009). A non-technical overview of GBGM will be discussed in Chapter 5 (Buitenweg et al., 2012; Hedden & Gabrieli, 2004; Jolles & Crone, 2012; Karbach & Schubert, 2013; Lövdén et al., 2012; Martin et al., 2011; Nagin & Odgers, 2010; Terrera et al., 2010).

**Baseline Characteristics as Predictors of Cognitive Training Responsiveness**

Given the heterogeneity of performance in cognitive training outcomes, it is also important to identify predictors of these inter-individual differences in cognitive outcomes (Nagin & Odgers, 2010; Tucker-Drob & Salthouse, 2011). Individual baseline characteristics - age, sex and proxies for CR (such as indices of intelligence and education) - are useful when investigating heterogeneous training outcomes. As noted in Chapter 3, these characteristics represent inherent cognitive capacity and lifetime experiences. They may influence an individual's receptiveness to intervention (Baltes, 1987; Boron et al., 2007a; Katzman, 1993). This is important when evaluating multi-domain programs in particular, as individuals may respond differently to the different elements of the training program (Martin et al., 2011). Exploration of individual baseline characteristics also leads to a more comprehensive exploration of the 'Use it or lose it' hypothesis and associated theories. Specifically, we can more precisely investigate to whom these the theories apply.

Identifying specific predictors of performance heterogeneity is also of practical use, as it allows individualised assignment of older adults to current treatment paradigms to maximise the training effect (*e.g.*, Baldi et al., 1996; Hastings & West, 2009; Hill et al., 1995; Kliegl et al., 1990; Rebok et al., 2007; Verhaeghen et al., 1992; West & Hastings, 2011). Prediction of inter-individual differences in cognitive training responsiveness may also encourage modifications of existing training programs or the development of alternative interventions or designs for those who show less benefit from existing training programs (Raz, 2009).[20] For example, more individually tailored approaches could be taken within these programs, such as varying the complexity of the tasks during training, individualised goal setting, consideration of motivation and arousal issues, as well as feedback and task variability (*cf.* Green & Bavelier, 2003; Martin et al., 2011). Supplementary sessions (e.g., pre-training or booster sessions) may also bolster cognitive change in those demonstrating less benefit from training (*e.g.*, Brooks, Friedman, Pearman, Gray, & Yesavage, 1999a; McKitrick et al., 1999). Overall, individually targeted training programs increase the cost effectiveness of training and may encourage government and industry investment

[20] Ideally cognitive training effects should be maintained without booster training (or with only minimal formal booster activity; Raz, 2009). As such, this review will only focus on studies without booster sessions.

(Fairchild et al., 2013; Hertzog et al., 2008). This is of particular importance when the costs of ageing are set to grow substantially with an ageing population (Papp et al., 2009). Thus, studies need to focus on such predictors.

The evidence from studies examining the effect of age, sex and CR proxies on training responsiveness is mixed. The use of conventional statistics may contribute to the ambiguous results. Investigation into whether inter-individual baseline characteristic differences are related to cognitive performance gains following training is in its infancy (*e.g.*, Park et al., 2007; Tucker-Drob & Salthouse, 2011; West & Hastings, 2011). Like the bulk of the literature, when predictors of inter-individual differences are investigated, the majority of studies focus on the cognitive domain of memory and few examine longitudinal effects. Thus, the current understanding of the influence of age, sex and CR proxies on the impact of training outcomes is unclear. The limited current understanding of the influence of these baseline characteristics is discussed below. The few studies that consider both inter-individual differences and longitudinal trajectories are highlighted.

**Age.** Age was the first characteristic explored to define individuals who demonstrate improvement following cognitive training. Age is also considered one of the most influential factors in determining general response to cognitive training in both animal and human studies. Hence it is the most commonly reported characteristic. However, there is mixed evidence from population-based studies as to whether increasing age affects training outcomes (Baltes et al., 2006a; Carey, 2007; Hertzog et al., 2008; Hertzog et al., 2009; Jessberger & Gage, 2008; Noack et al., 2009; Schaie & Willis, 2010).

In line with the investigation of age in the context of testing the upper limits of cognitive performance overall, the training literature often emphasises the negative relationship between age and level of plasticity or reserve-building mechanisms (Baltes et al., 2006b; Brown et al., 2003; Carey, 2007; Hertzog et al., 2008; Jessberger & Gage, 2008; Jones et al., 2006; Kempermann, Kuhn, & Gage, 1998; Lustig et al., 2009; Noack et al., 2009; Schaie & Willis, 2010). That is, gains are often considered less common with advancing age. It is suggested that cognitive decline predominates with the progression of ARCD, beyond the benefits of training effects (Hertzog et al., 2008; Whitlock, McLaughlin, & Allaire, 2012).

Some of the literature examining memory performance suggests that younger adults benefit more from training than older adults (Bissig & Lustig, 2007; Boron, Willis, & Schaie, 2007b; Brooks et al., 1999a; Gehring et al., 2011; Hill et al., 1995; Lustig et al., 2009; Sheikh, Hill, & Yesavage, 1986; Singer, Lindenberger, & Baltes, 2003; Verhaeghen et al., 1992; Verhaeghen & Marcoen, 1996; Yesavage, Sheikh, Friedman, & Tanke, 1990; Zelinski et al., 2008). Singer and colleagues (2003) observed that within samples of older persons, the individuals aged in their 60s and 70s exhibited larger training-related gains in mnemonic training than those even older (≥80 years). Cohort comparisons for the SLS suggested that training benefits were least readily accomplished for the oldest-old in the cohort (Lovelace & Twohig, 1990; Willis, 1989). Similarly, from meta-analyses and their own research, Verhaeghen and colleagues (1992) reported that whilst both younger and older adults can benefit from mnemonic training, older trainees gain less than younger trainees (Verhaeghen & Marcoen, 1996; Verhaeghen et al., 1992). Thus, these studies suggest that with older age, the capacity for improvement is reduced, such that training cannot compensate for the overall loss of performance.

In contrast, no significant effect of age has been demonstrated (Ball et al., 2002; Denney & Heidrich, 1990; Hallett, 2001; Hertzog et al., 2009; Karlsson et al., 1989; Kramer & Willis, 2003; Park et al. 2007; Schaie & Willis, 1986; Wahlin et al., 1993; Willis & Nesselroade, 1990). In least in some healthy older-old adults, the plasticity mechanisms appear to be preserved, allowing equal acquisition of trained skills of older-old versus younger-age individuals. Also, Salthouse (2006) states that some relatively young older adults may demonstrate maximal levels of plasticity, as evidenced by peak cognitive performance for their age. That is, because they have less room for plasticity-induced improvement than those relatively older, no effects of age are evident. Other instances of ceiling effects of age on cognitive functioning following training have been demonstrated in the literature (*e.g.*, Fairchild et al., 2013; Lövdén et al., 2012).

As previously noted, the use of conventional statistics may contribute to the mixed results with regard to the predictive effect of age. Much of the above-cited literature used group-based, correlational data and many were pre-post comparisons (Ball et al., 2002; Herlitz et al., 1997; Martin et al., 2011). Studies exploring predictors of inter-individual responsiveness to training effects, particularly longitudinally, are scarce. However, the longitudinal studies discussed below show further mixed evidence of the effect of age on training gains.

McKitrick and colleagues (1999) and Fairchild and colleagues (2013) examined predictors of *inter-individual responsiveness* following training in a series of nested studies involving cognitively healthy older adults (aged 55–83 years). McKitrick and colleagues (1999; $n$ = 224) investigated immediate effects and found that successful participants on a delayed word recall task were more likely to be younger, but age was not a significant predictor of short-term response to a name recall task.

In contrast, Fairchild and colleagues (2013; $n$ = 120) found that age was not a significant predictor of longitudinal success (one year post training), yet younger age was a predictor of long-term treatment response on a name-face recall. Those who were younger than 65 years of age scored the best response (70.7% versus 37.8% of individuals 65 years or older).

Together, the findings from these two studies highlight that the type of task, and the time frame of training effects (*e.g.*, immediate or longer-term effects), adds to ambiguity in the literature. They show that age needs to be further investigated as a predictor when looking at inter-individual differences in *longitudinal* training effects across multiple domains.

Studies investigating cognitive *trajectories* following training are also scarce, particularly in the ageing context. The wider rehabilitation literature investigating traumatic and acquired brain injured populations provides some evidence of the effect of age on performance trajectories; however results are again inconsistent (Green et al., 2008a; Zwaagstra, Schmidt, & Vanier, 1996). These brain injuries could be analogous to the 'damage' associated with ARCD.[21]

Some studies suggest that younger age moderates improvement in cognitive trajectories across time (Gehring et al., 2011; Green et al., 2008a; Zwaagstra et al., 1996). Gehring and colleagues (2011) conducted cognitive training with 366 adult patients with glioma (cancer of the glial cells; age $M$ = 41.8, $SD$ 9.5). Using multi-level modelling, they demonstrated that six months following cognitive rehabilitation younger adult patients were more likely to demonstrate a benefit compared with older patients.

[21] Although as noted ARCD is conceptualised as a normal process in this review.

In contrast, age has not been found to be a predictor of performance trajectories in other rehabilitation studies (Chu et al., 2007; Fairchild et al., 2013). Chu and colleagues (2007) also used a multi-level modelling approach to measure recovery in new learning and memory following traumatic brain injury. They found that factors such as age only predicted the level of cognitive outcome, rather than influencing the recovery trajectory.

In sum, the evidence for the impact of age on training effects is mixed. Whilst much of the memory training literature shows a negative effect of age on performance when using group level, pre-post analysis this is not always the case. Some studies report no significant differences in training gains for different age groups. Furthermore, the effect of age may differ, depending on the cognitive measure used and the length of follow-up (*e.g.,* Fairchild et al., 2013; McKitrick et al., 1999). Importantly such mixed results even occur when considering longitudinal change and cognitive trajectories (the latter at least in the broader rehabilitation context). It is therefore evident that age should be further explored as a predictor, in the context of inter-individual differences in longitudinal cognitive performance trajectories in older adults following training (Goldin, 1998; Jones et al., 2011; Reynolds et al., 2002; Stern, 2002; Schaie et al., 2005; Valenzuela & Sachdev, 2006a; Valenzuela & Sachdev, 2006b; Yaffe et al., 2009).

**Education.** The influence of education on cognitive performance in older adults has been discussed and debated for decades and its effects, like age, are unclear (*e.g.*, Dunlosky & Hertzog, 1998a; Sullivan, 1964; West & Tomer, 1989). In the training literature, education has been conceptualised as a representation of, and/or as influenced by, plasticity (Hill & Bäckman, 2000; Lövdén et al., 2010). Education is more commonly discussed as a proxy for CR, as noted in Chapter 2 (Goldin, 1998;

Jones et al., 2011; Reynolds et al., 2002; Stern, 2002; Schaie et al., 2005; Valenzuela & Sachdev, 2006a; Valenzuela & Sachdev, 2006b; Yaffe et al., 2009). Furthermore, greater education has been described as a measure of increased learning ability[22] (Garrett et al., 2012; Stern, 2002; Stern, 2009). These conceptualisations emphasise that higher educational attainment enables improved cognitive performance following training. That is, individuals with higher levels of education have greater capacity for plasticity, more efficient processing, and have a greater learning ability to assist in transfer of information from temporary to permanent storage within the brain (Bagwell & West, 2008; Craik, 1983; Hill et al., 1995; West & Hastings, 2011).

However, of the limited evidence available in relation to the predictive value of education on training gains, the results are mixed. Some evidence shows higher levels of education are associated with improved cognitive performance following training (Bagwell & West, 2008; Garrett et al., 2012; Gross et al., 2011; Hill et al., 1995; Hill & Bäckman, 2000; Langbaum et al., 2009; Lövdén et al.; Stern et al., 1994; West & Hastings, 2011; Zelinski et al., 2008).

In the cognitive rehabilitation literature, Gehring and colleagues' (2011) previously noted study of 366 patients with glioma showed that higher levels of education were associated with reliable improvement across time following training, at least in middle-aged adults.[23] Gehring and colleagues concluded that education may not only buffer the outcome of brain damage and influence the initial cognitive status after

[22] Education has also been described as a measure of intelligence (Garrett et al., 2012; Stern, 2002; Stern, 2009). Whilst it is acknowledged that education and intelligence are tightly linked (*e.g.*, through the 'multiplier effect', noted in Chapter 3), this review discusses and considers education and intelligence as separate potential predictors of cognitive training effects.

[23] Again it is acknowledged that cognitive training is distinct from cognitive, rehabilitation, although here the cognitive rehabilitation literature provides a conceptual analogy of the benefit of intervention.

injury (as CR theory would suggest), but it may also play a supporting role in restitution of function or functional reorganisation during recovery (in line with its representation of plasticity). That said, some rehabilitation studies show that education does not moderate improvement in cognitive functioning for patients with traumatic brain injury (Green et al., 2008a; Zwaagstra et al., 1996).

Of the few training studies exploring education in the cognitive training literature, Hill and colleagues (1995) showed the benefits of higher levels of education on training outcomes in older adults ($n$ = 253). Education was investigated as a mediator in post-training episodic memory improvement, with a greater number of years at school being positively related to performance improvement, even after controlling for the effects of age.

In contrast, education has also been shown *not* to have an effect on training outcomes (*e.g.*, Green et al., 2008a; Verhaeghen et al., 1992; Zwaagstra et al., 1996). In Verhaeghen and colleagues' (1992) meta-analysis, education was explored as a possible important mediator variable. However, education failed to yield a significant between-groups difference. Therefore the literature examining the influence of education on memory performance gains is equivocal, similar to the ambiguous influence of age on training outcomes.

Few studies look beyond the memory domain to EF when examining the effect of education. Those that do also show mixed effects (Boron et al., 2007a; Schaie & Willis, 1986; Willis & Nesselroade, 1990). Boron and colleagues (2007a) examined effects cognitive training on reasoning ability and gains in inductive reasoning performance ($n$ = 335) in the Seattle Longitudinal Study (SLS). They found that individuals with lower education showed a greater magnitude of change from pre-test to post-test in accuracy of inductive reasoning performance following training than higher functioning individuals. In contrast, both Schaie & Willis (1986) and Willis & Nesselroade (1990) found no effect of education on inductive reasoning and spatial orientation performance.

Thus, overall, the effect of education on memory and EF following training in older adults is unclear (Langbaum et al., 2009). Despite the extensive discussion and debate around the influence of education on training, quality research specifically predicting training outcomes based on education level is sparse. It is therefore evident that education should be further explored as a predictor of training outcomes, and also in the context of inter-individual differences in cognitive performance trajectories.

**Indices of intelligence.** There has been long standing debate with regard to the influence of intelligence on cognitive training outcomes (Garlick, 2002; Lövdén et al., 2010). Higher levels have most often been mooted as supporting learning conditions in cognitive training in a similar way to education (Bagwell & West, 2008; Hill et al., 1995; West & Hastings, 2011). Proponents see higher indices of intelligence equating to higher training capacity and efficiency to learn and implement training protocols (Garlick, 2002; Lövdén et al., 2010). As previously noted in Chapter 3, indices of intelligence, are seen as a proxy for CR (Buckner et al., 2004; Christensen et al., 2007; Stern, 2007; Valenzuela & Sachdev, 2006a; Valenzuela & Sachdev, 2006b), with those individuals with higher IQ being better able to build on their existing scaffold of cognitive mechanisms and processes through plasticity (Bagwell & West, 2008; Gehring et al., 2011; Green et al., 2008b; Hertzog et al., 2009; Hill et al., 1995; Hunt, 1978; West & Hastings, 2011). Thus, as with education in the rehabilitation context, individuals with greater levels of intelligence not only have a greater initial intrinsic buffer against ARCD – *i.e.*, greater reserve against 'damage' – but also have a greater capacity for restitution of function or functional reorganisation to boost cognitive performance as damage occurs (Gehring et al., 2011; Green et al., 2008b).

In addition to plasticity, indices of intelligence represent cognitive flexibility (Lövdén et al., 2012). Those with higher IQ may be better able to use relevant information structures or knowledge and make the necessary conscious efforts to select and implement strategies – such as those taught in cognitive training programs – in order to execute a behavioural outcome, for example, a mnemonic strategy (Carey, 2007; Hill et al., 2000; Jones et al., 2006; Lövdén et al., 2010; Noack et al., 2009).

However, the evidence for the impact of indices of intelligence on training outcomes is varied (Carter, 2002; Hill et al., 2000; Neils-Strunjas, Krikorian, Shidler, & Likoy, 2001; Sullivan, 1964; Yesavage et al., 1988; Zelinski et al., 2008). Some studies support the influence of intelligence (*e.g.*, psychometric *g* and Wechsler Adult Intelligence Scale [WAIS], crystallised intelligences[24]) as aiding learning and memory training gains (Bretz & Thompsett, 1992; Carter, 2002; Pervin, 1978; Ree & Earles, 1991; Wexley, 1984; Yesavage et al., 1988). In a small study Yesavage and colleagues (1988) found a correlation between participants with greater verbal ability (Wechsler Adult Intelligence Scale; WAIS vocabulary subscale score) and a greater benefit from a memory training paradigm combining mnemonics and verbal elaboration techniques ($n$ = 40 older adults). Lövdén and colleagues (2010) further describe individuals' greater levels of knowledge as creating more cognitive flexibility, with more methods to execute a behavioural outcome, such as strategies taught in cognitive training programs.

[24] Defined as constructs which represent knowledge that an individual has obtained through experience, such as vocabulary (Brickman & Stern, 2009).

In contrast, some studies showed no link between indices of intelligence and training outcomes (Neils-Strunjas et al., 2001; Zelinski et al., 2008). Neils-Strunjas and colleagues (2001) conducted a small memory training program study with 50 older adults (age *M* = 72.2 *SD* = 7.4). Participants were shown videos and asked to learn the first and last names of the 20 actors (*i.e.*, 40 names in total). They found no correlation between recall performance and general mental ability.

When considering the impact of premorbid intelligence in studies examining performance trajectories, the rehabilitation literature also shows no effect (Chu et al., 2007; Green et al., 2008a). Both studies by Chu et al. and Green et al. demonstrated that pre-morbid intelligence only predicted the initial level of cognitive performance. Thus again, like age and education, the evidence for the impact of intelligence thus far is ambiguous and overall further research is required.

**Sex.** There have been very few explorations of the impact of sex on cognitive training outcomes in the ageing literature. Of the few studies, the findings are inconsistent (Gross & Rebok, 2011; Gross et al., 2012; Herlitz et al., 1997; Schaie, 1994; Schaie & Willis, 1986).

Boron and colleagues (2007a) investigated gains in inductive reasoning performance following reasoning ability training in older participants of the SLS. The authors found that gains were seen in *females*, not males. Similarly, two studies by Schaie and colleagues (1986; 1994) showed that females demonstrated greater gains on spatial orientation tasks from spatial orientation training. However, the authors showed *males* benefited more from inductive reasoning training. McKelvie and colleagues (1993) reported sex differences in recognition memory for faces versus cars. Females recognised more female and children’s faces than males, yet males performed at a higher level of recognition of male faces and cars. Thus there are sex differences which may be related to the type of task.

Furthermore, the extent of the gains may be influenced by baseline performance (*i.e.*, cognitive performance prior to training). Schaie and colleagues (1986; 1994) noted that females had lower baseline spatial orientation levels compared with males. Once this effect was removed, there was no sex difference for that ability. Indeed sex may an important impact on individual differences in cognitive function overall (*e.g.*, Herlitz et al., 1997; Hirshson, 2010; Jones et al., 2005; Kaplan, 2004; Larrabee & Crook, 1993; Meinz & Salthouse, 1998; Terrera et al., 2010; van Hooren et al., 2007; West, Crook, & Barron, 1992; Wiederholt et al., 1993; Zelinski et al., 1993; Zelinski & Stewart, 1998). Males have been found to outperform females on spatial tasks in general (Schaie & Hertzog, 1986). Others suggest females outperform males on VM (Colsher & Wallace, 1991; Herlitz et al., 1997; van Hooren et al, 2007; Hultsch, Hertzog, & Dixon, 1990; Meinz & Salthouse, 1998; West et al., 1992; Zelinski et al., 1993).

A number of explanations have been posited for sex differences across cognitive domains. Authors have suggested that sex differences may be a result of task demands, which differentially engage male or female interest, familiarity or motivation (McKelvie et al., 1993). Biological differences between males and females, such as brain asymmetry, declines in brain volume or the influence of sex hormones may contribute (Gur et al., 1991; Lezak et al., 2004; Yaffe, Lui, Zmuda, & Cauley, 2002). In addition, others note differences in life expectancy (Hooyman & Kiyak, 2002), which, in turn is linked with the theory that males may be closer to death or "terminal decline"[25] than females (Boron et al., 2007a) and thus lower performance at baseline and greater room to improve. Similarly, unhealthy behaviours are more frequent in males (Artaud et al., 2013) and that, in later life, males are significantly less engaged in more active cognitive lifestyles than females (Valenzuela et al., 2013). According to the 'Use it or lose it' and associated theories, less engagement may contribute decreased performance at baseline.

A higher cognitive performance level may, however, leave less room for improvement on some cognitive measures (*i.e.*, a ceiling effect), as was noted above when discussing the influence of age, education and premorbid IQ on training gains. Thus, if females have higher existing levels of memory, for example, males will be able to demonstrate greater improvement following training.

Whist there is support in the literature for sex differences in performance following training, there is, however, evidence in the training literature that sex has no significant relationship to the size of performance gains following training (Gehring et al., 2011; Hill et al., 1995; Hill et al., 2000; Kaplan, 2004; Medalia & Richardson, 2005; Zelinski et al., 2008). For example, Hill and colleagues (1995) explored demographic characteristics on episodic recall tasks ($n$ = 253). The authors found that sex did not influence the magnitude of memory performance gains.

Thus, overall, the evidence is ambiguous with regard to the effect of sex on cognitive performance following training, although this may be in part a result of the different cognitive domains measured and initial performance. Therefore further research is necessary across a range of cognitive domains.

[25] A controversial concept, 'terminal decline' or 'terminal drop' is an increased rate in cognitive decline as one approaches death, often reported as beginning 3–6 years before death (Jarvik & Falek, 1963; Lieberman, 1965; Riegel & Riegel, 1972 in Hedden & Gabrieli, 2004; Wilson et al., 2007).

**Baseline Characteristics as Predictors of Interindividual Differences and Trajectories of Cognitive Performance Following Training**

As highlighted throughout this review, the lack of clarity in the literature with regard to the effect of training overall may also be the insufficient number of quality studies. Impacting this are the few studies utilising appropriate statistical techniques to consider the influence of baseline characteristics on inter-individual differences in training performance trajectories (Gross et al., 2011; Herlitz et al., 1997; Martin et al., 2011; Raz, 2009). Recently, growth modelling techniques have been utilised. Whilst some of these studies in the wider rehabilitation context have been noted, two important studies investigating participant baseline characteristics will now be discussed (Gross et al., 2012; Langbaum et al., 2009). These studies assess multiple baseline characteristic predictors and perhaps provide the most accurate form of investigation of training effects in the literature thus far.

Gross and colleagues (2012) used LGMs of performance to investigate age, sex and years of education as predictors of performance trajectories in the memory-trained group ($n$ = 1,401) from the ACTIVE study. Specifically, they reported on initial recall (trial 1) and learning across trials (sum of words recalled across trials) in the RAVLT and Hopkins Verbal Learning Test (HVLT). When considering immediate training effects, Gross and colleagues (2012) found that sex (*i.e.*, being female) was associated with *less* immediate post-training-related improvement in initial recall on VM tasks (the RAVLT and HVLT). Age and education were not predictive of immediate training effects in initial recall in the two memory measures. Longitudinal effects of training on initial recall performance were also assessed after five years. The authors found that only *age* was a significant predictor for the RAVLT and HVLT trial 1 performance: being younger predicted better performance (*i.e.*, slower decline from initial recall). Education and sex were not longitudinal performance predictors. Age, sex and education did not predict immediate post-training-related improvements for either the RAVLT or the HVLT on learning performance. Long-term learning change across the five-year follow-up was also not predicted by these baseline characteristics.

The study by Gross and colleagues (2012) highlights that with different outcome variables and different lengths of long-term follow up, baseline characteristics have a different effect on cognitive performance trajectories following training. For example, age was only predictive of initial recall and not of learning curve performance. This is important to take into account when considering the longitudinal effects of cognitive training on performance trajectories.

Langbaum and colleagues (2009) also examined age, sex and education of participants in the memory training arm of the ACTIVE study (*n* = 703). The authors investigated differential performance trajectories of participants based on training responsiveness to a composite of memory tasks. Using latent class analysis, they identified three distinct sub-groups (or classes). The first sub-group (labelled the 'HVLT class'; *n* = 123) consisted of individuals defined as having a high conditional probability of responding to training on the HVLT total score as well as a moderate conditional probability of responding on the HVLT discrimination and the Rivermead tests administered. The second sub-group (the 'RAVLT class'; *n* = 210; *i.e.*, those who had a higher conditional probability of responding on trial 1 of the RAVLT) benefited less across the various cognitive outcome tasks administered than the HVLT class. Individuals had a high conditional probability of responding to the RAVLT measure and moderate-to-low conditional probabilities on the other measures. Finally, the third group was defined as being the "low-level response class" (p.15; *n* = 271), including individuals showing a low-to-moderate conditional probability of responding to training as indicated by the total score from each memory measure. This class also consisted of those who showed either a decline in performance or no distinct pattern of responsiveness to training.

Following identification of these classes, analysis revealed that education and age were predictive of these distinct response patterns. Higher education was predictive of being a member of one of the higher-performing classes compared with the “low-level response class”. Interestingly, younger age was also predictive of being in the lower performing class compared with the highest performance class (the RAVLT class), which is not consistent with other indications of a positive effect of younger age on performance gains.

Both these studies demonstrate a number of key points. Firstly, the studies show that baseline characteristics can effectively be utilised to identify the inter-individual differences in cognitive performance trajectories. In addition, they show that different individuals profit differently on different cognitive tasks; thus is it important to consider different cognitive outcome measures when assessing training outcomes. Finally, the studies also highlight the utility of more sophisticated statistical approaches to identify predictors of heterogeneous training responsiveness.

## Summary

In sum, investigation into the efficacy of cognitive training is ambiguous. There is a lack of demonstration of generalisability and long-term efficacy of cognitive training. Also, the effect of age, education, indices of intelligence and sex differences in cognitive performance following training shows mixed results. Whilst there is some plausibility with regard to younger age, higher levels of intelligence and education and sex predicting training gains, this is not always the case. Methodological limitations, including statistical techniques, contribute to this ambiguity. Only a handful of studies adequately considered baseline characteristics as predictors, inter-individual differences and/or trajectories of change. Task type may influence the effect of baseline features, for example, types of memory. Cognitive domains other than memory, such as executive functions, have not been thoroughly explored. It is clear further exploration of distinct cognitive trajectories across time is warranted through the use of new and more sophisticated statistical techniques.

**LITERATURE REVIEW**

# Chapter 5

# Group-based Growth Modelling and Related Techniques

Group-based growth modelling (GBGM) is a relatively new statistical technique that has fundamentally altered how we conceptualise and study change across time (Duncan et al., 2000). Group-based growth modelling is important in the context of assessing cognitive changes following training in older adults, given the issues with current methodological approaches outlined in Chapter 4. Given the relative novelty of the technique, the present chapter outlines and discusses its statistical standing. The discussion will be largely from a non-technical perspective. The chapter also highlights the key empirical components, and the common and recommended analytic procedures by which the models are produced and selected. In addition, limitations to these techniques are elucidated. This leads to an explanation for the specific group-based modelling (GBM) approach utilised in the current study – *generalised growth mixture modelling* (GGMM) – used to evaluate older adults' cognitive trajectories following training. To further explain the concepts discussed, a worked example based on the current study is used throughout the chapter, showing cognitive performance tested across four time points (baseline, three-, six- and 12-months post-baseline). This is schematically represented in Figure 2.

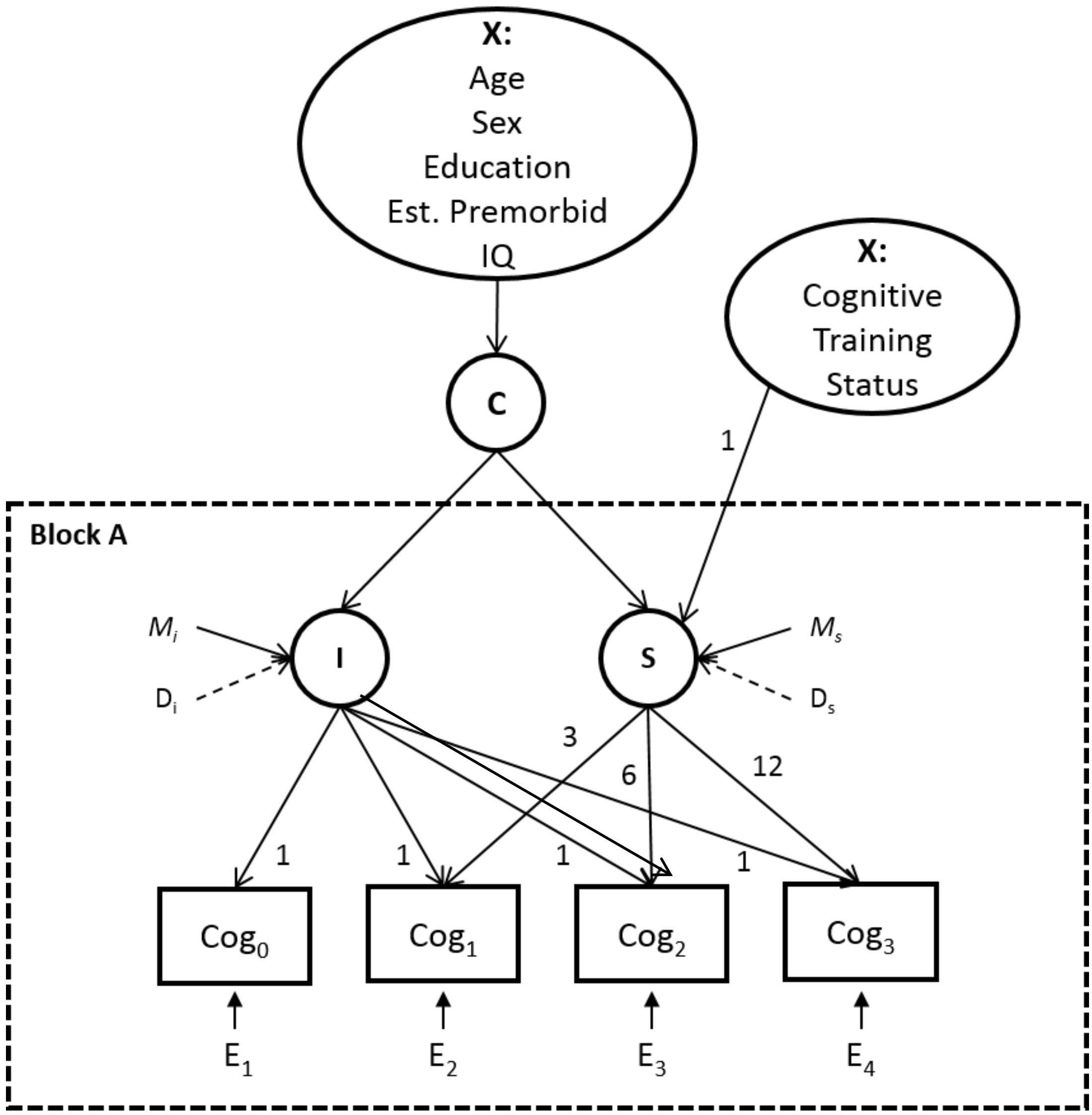


*Figure 2.* A linear group-based growth model schematic.

*Notes:* Observed, continuous variables $Cog_0$, $Cog_1$, $Cog_2$, $Cog_3$ represent repeated measures from baseline across the 3-, 6- and 12-month follow ups, respectively. $Cog_0$ is the baseline level of performance. Cognitive training (for those in the experimental group) begins at $Cog_1$. The covariate (X) is the predictor. The path from the Cognitive Training Status covariate to $Cog_0$ is not demonstrated, because this is fixed at 0, given it is the baseline (Acock, 2005). This also applies to the path from the Cognitive Training Status covariate the intercept. E = measurement error. *M* = mean, *i.e.*, conditioned means $M_i$ = mean intercept and $M_s$ = mean slope. D = the variance, *i.e.*, $D_i$: intercept variance and $D_s$: intercept slope. The intercept is identified by the constant loadings of 1 going to each Cog score. Block A (indicated by the dotted square) represents the initial, latent growth model (LGM) without any covariates.

Group-based growth modelling has been used in various disciplines to examine change, including medicine, education, criminology, the social and behavioural sciences, and psychology. More recently these techniques have been used in treatment studies, in which temporal responses to intervention are investigated (*e.g.*, Hix-Small, Duncan, Duncan, & Okut, 2004; Muthén et al., 2002; Muthén & Muthén, 2002; Muthén & Shedden, 1999; Rodriguez, Moss, & Audrain-McGovern, 2005; Stulz, Gallop, Lutz, Wrenn, & Crits-Christoph, 2010; Sterba, Prinstein, & Cox, 2007). Adoption of this method in the ageing and cognitive training literature is pertinent and necessary although use is currently in its infancy (*e.g.*, Leoutsakos et al., 2012; Terrera et al., 2010; West & Hastings, 2011).

There are numerous statistical conceptualisations of GBGM. A simple conceptualisation highlights both its variable-centred and, as previously noted, person-centred approach (Jung & Wickrama, 2008; *cf.* Muthén & Muthén, 2000). The variable-centred component of the model describes relationships amongst dependent and independent variables. It can be considered to include features of regression, latent variable modelling and structural equation modelling (SEM) frameworks. Importantly, this component can also include the investigation of outcome predictors, a strength that is further discussed later in this chapter.

The person-centred component focuses on the relationships between individuals. Person-centred methodologies, perhaps less familiar in statistics, include cluster analysis, taxometrics and finite mixture modelling, whereby inter-individual differences are considered and the population of interest is considered to be heterogeneous (Jung & Wickrama, 2008). Heterogeneity is considered by grouping individuals into a finite number of categorical latent classes (or latent variables) defined by their developmental trajectories (Duncan et al., 1999; Duncan, Duncan, & Strycker, 2000; Muthén, 2004; Nagin & Odgers, 2010). That is, individuals with similar temporal performance trajectories are allocated to the same class (Muthén et al., 2002). Inter-individual differences are considered across these classes (Muthén &

Muthén, 2000). Membership of these classes is not known, but is inferred from the data. Representation of different classes is a feature distinguishing GBGM from more traditional, LGM (Meredith & Tisak, 1984; Meredith & Tisak, 1990; Rao, 1958; Scher et al., 1960; Tucker, 1958). Latent growth modelling, however, provides a useful conceptual and practical starting point for understanding and applying GBGM.

## Latent Growth Modelling

Conventional, LGM, also known as *Latent Growth Curve Modelling* (LGCM), represents the most simple form of trajectory modelling (Leoutsakos et al., 2012; Meredith & Tisak, 1990; Nagin & Odgers, 2010; Rao, 1958; Scher et al., 1960; Tucker, 1958). Latent growth modelling utilises some of the core components of GBGM. Latent growth modelling is also considered a good base from which to build more complex, conditional and group-based models, particularly for beginners to GBGM techniques (Acock, 2005; Cuijpers, van Lier, van Straten, & Donker, 2005; Jung & Wickrama, 2008; Li, Duncan, Duncan & Acock, 2001; Muthén et al., 2002; Stulz et al., 2010). Latent growth modelling estimates the *common average growth across time*, encompassing all performance trajectories and follows the simplest possibility as the null hypothesis: that a single, growth curve model can characterise performance across time. It incorporates an implicit assumption that the data are from a single, homogenous population.

### Model Parameters and Procedures in Latent Growth Modelling

The shape of the model is constructed using a number of parameters (Acock, 2005; Duncan et al., 2002; Jung & Wickrama, 2008; Muthén, 2004). These parameters will be discussed below and are demonstrated in Figure 2.

**Outcome measure.** The outcome measure is referred to in the growth modelling literature using a number of different terms based on the underlying statistical conceptualisation and application (*e.g.,* statistical program in which it is being applied). Terms include the continuous latent variable, proximal outcomes measure, and the observed or measured variable, and/or it is referred to by the letter *y* (Acock, 2005). The continuous outcome measure in Figure 2 is represented by the squares $Cog_{x,}$ representing each measurement of cognitive performance at each time point.

Outcome variable parameters are components of the outcome measures. Outcome parameters include residual variances, time-specific and measurement error variation (also known as error variance). Outcome parameters are represented in the literature by a variety of symbols including ε, δ and Ψ (Li & Acock, 1999). In the worked example in Figure 2, E is used. Measurement errors are a strength of the modelling technique, because they demonstrate the uniqueness associated with measurement of an observed variable. Measurement errors thereby represent the assumption that the measurements at each time point are not perfectly reliable (Li & Acock, 1999).

**Intercept and slope.** Performance trajectories are related to time through a regression function using key parameters referred to as the intercept (I) and slope (S). They are also termed continuous latent variables, growth factors or within-population growth parameters and/or represented mathematically by η or F (Acock, 2005; Muthén, 2004). Both the intercept and slope can be considered constants: α or Const (Acock, 2005; Muthén, 2004). The *intercept* represents the estimated mean initial level of performance; that is, where the estimated growth curve begins. It is represented by $M_i$ in Figure 2. Measurement invariance is represented here by constant loadings of 1 to each testing session (Muthén, 2005).

The *slope* represents the estimated mean rate of growth ($M_s$ in Figure 2). Like the intercept, the slope is identified by fixing the values of the paths to the primary outcome variable at each testing session. The time elapsed between test sessions is specified (*e.g.*, if each testing session is equidistant from each other, then 0, 1, 2, 3, *etc.*). The current worked example in Figure 2 shows a non-equidistant time-frame. Paths are fixed with the numbers 3, 6 and 12 (*i.e.*, the time scores representing the number of months post-baseline). No line is demonstrated from S to $Cog_0$ since this is fixed at 0 (*e.g.*, Acock, 2005). When the slope growth factor mean is significant (*i.e.*, it is significantly different from zero) the model is showing significant development over time, on average (Acock, 2005).

The slope can be described further by other parameters, including whether it is linear or nonlinear, *e.g.*, quadratic or piecewise linear (Li & Acock, 1999). At least three outcome variable measurements are required to estimate a linear trend to estimate both the *amount* and the *shape* (*i.e.*, *rate*) of the trajectory (Duncan et al., 2000). Four measurement points allow estimation of a quadratic trend, by changing the scale of the Y-axis with the addition of a quadratic term (Acock, 2005; Li & Acock, 1999). The worked example and current study is represented by a linear slope. Linear slopes are considered simpler models, better suited when there are relatively smaller sample sizes (Acock, 2005). This is presented in the Block A in Figure 2.

***Intercept and slope variance.*** Latent variation across individuals can also be modelled as an additional model specification. Therefore LGM can be used to represent some heterogeneity in the model depending on if heterogeneity is expected, and depending on the research questions posed (Jung & Wickrama, 2008).[26] That is, in addition to the group average intercept or slope level, each individual's distinct performance can be demonstrated by the variance around the intercept and slope. When allowing for variance in LGM the model produced is also called a multi-level, random effects model – also known as a mixed model, or two-level random coefficients model (Jung & Wickrama, 2008; Li & Acock, 1999; Muthén, 2004; Nagin & Odgers, 2010).[27] This method is also described as the residual associated with the prediction of the latent variables (Li & Acock, 1999). The variances are critical when exploring more complex models with *covariates* (also called predictors,[28] *e.g.*, training effects and baseline characteristics), to determine whether they explain the latent variables. For example, why some individuals have a steeper or less steep growth rate than the average (Acock, 2005). This is further discussed below in the context of conditional LGMs.

Figure 2 demonstrates the latent variance around both the intercept and slope growth factors ($D_{i,}$ and $D_{s,}$ respectively). The intercept variance represents the average difference between individuals' intercepts. The slope variance represents the average differences in individual growth curves. Large estimated values indicate initial performance levels or rates of change that vary widely. Small variances demonstrate that the group is more homogeneous. Statistical tests represent the significance of the variance. When growth factor variance is significant, this indicates heterogeneity in individuals' starting points or trajectories. If the growth factor variance is not significant, this represents relative homogeneity in these parameters. In the current

---

[26] Variation is represented by the random effects around the intercept and slope and is described as the residual associated with the prediction of the latent variables (Li & Acock, 1999).

[27] When conducting a traditional LGM the variances are fixed to zero

[28] The term 'covariate' will predominantly be used in this statistical chapter. Covariates will be referred to as predictors when conclusions drawn from model data are emphasised.

worked example, significance would equate to heterogeneity of levels of performance on the cognitive test at baseline, and/or different changes in this performance across the 12 month follow-up period, respectively. It is common to initially conduct an analysis without estimating the variance, *i.e.*, the variance is fixed at zero (*e.g.*, Cuijpers et al., 2005; Nagin, 1999; Stulz et al., 2010). This can then be compared to models in which variance was estimated (*e.g.*, Muthén, 2002).

There are some limitations, however, to utilising the LGM to represent heterogeneity. Allowing for variance around the growth curve increases model complexity and, as such, inadmissible models are commonly produced, due to non-convergence, local solutions and/or small or negative variances (Muthén, 2005).[29] Furthermore, as noted, LGM assumes there is a common average growth across time, with individuals differing to a lesser or greater extent around this trajectory. It can be unrealistic to assume that a single trajectory best represents the data, particularly with regard to inter-individual differences in cognitive ageing trajectories becoming more apparent in the ageing literature. For example, the presence of small extreme groups may dominate the patterns for an entire sample, mask heterogeneous developmental pathways and/or bias the identification of distinct trajectories and ultimately obscure identification of causal dynamics (Duncan et al., 2002; Muthén, 2002). Heterogeneity of performance responsiveness seen in the training context is one such area where this may occur. Conducting LGM alone may bias identification of distinct patterns of person-centred trajectories, and thus obscure the causal dynamics leading to diverse outcomes across classes (Connell & Frye, 2006; Duncan et al., 2002). Thus LGMs, like conventional statistics, may not appropriately model cognitive trajectories.

[29] Model non-convergence and local solutions are further discussed in the section discussing limitations and controversies of the GBGM below. Also see Hipp & Bauer (2006).

In addition the variance of the intercept and the slope can be *correlated*, further specifying individual trajectories (Li & Acock, 1999). This would identify, for example, whether an individual who starts at a low level of cognitive performance will grow more quickly that those who begin at a higher level. This has not been represented in Figure 2, as this procedure was not carried out in the current empirical study to decrease the chances of model unacceptability due to the increased model complexity.

**Separation of Models for Each Group**

Separate LGM analyses for the control and experimental groups, as well as for the entire study sample, are often conducted within experimental research. Each population can be regarded as different, with distinctive trajectories (Muthén & Muthén, 2002). For example, it is important to identify a separate trajectory from the control data, representing normative growth, *i.e.*, without the effect of the intervention (Jöreskog & Sörbom, 1979; Muthén et al., 2002). Models for each group are particularly necessary when there is non-randomisation of the sample (as is the case in the current study).

**Conditional Latent Growth Modelling: The Inclusion of Covariates and/or Distal Outcomes**

The LGM model described above can be considered an unconditional model (Bryk & Raudenbush, 1992). Following the estimation of the growth trajectory, including specifications with the various parameters explained above, the unconditional LGM can be further extended to form a conditional model to predict growth/change or attempt to explain a variable that is influenced by the growth processes (Acock, 2005; Chen & Cohen, 2006; Nagin & Odgers, 2010). This can be achieved by including *covariates* (predictors; X) and/or a *distal outcome* (*U*), respectively (Acock, 2005; Duncan, Duncan, & Stryker, 2006; Muthén, 2004). These can therefore help explain the developmental trajectory and/or assess response to clinical interventions (Nagin & Odgers, 2010).

As in ordinary linear regression, covariates can be ‘time invariant’ and constant across the time points (‘fixed’ covariates). These are often measured at baseline, *e.g.*, treatment group, participant characteristics such as age, sex, education and estimated premorbid IQ in the current worked example and study. In contrast, covariates may be ‘time-varying’, thus their values change over the multiple assessments when they are considered. These covariates are either measured after the process has started or have a value that changes, *e.g.*, psychiatric disorder, hours of training (Acock, 2005; Chen & Cohen, 2006). In either case covariates are added to determine their association with the developmental course, that is, the extent to which they may account for a variance of the sample mean or linear trajectory (Chen & Cohen, 2006). A *distal outcome* included in the model demonstrates if certain consequences are influenced by the growth process (Huang et al., 2010; Muthén et al., 2002; Muthén, 2004). The inclusion of covariates and distal outcomes is recommended as a latter step in building LGM model complexity (Acock, 2005; Jung & Wickrama, 2008; Muthén et al., 2002).

Conditional LGM has been used in a number of studies (*e.g.*, Cuijpers et al., 2005; Duncan et al., 1999; MacCallum, Kim, Malarkey, & Kiecolt-Glaser, 1997; Muthén & Muthén, 1998–2010). In a study examining differential effects of psychological treatment of major depressive disorder, Cuijpers and colleagues (2005) utilised the conditional LGM technique prior to conducting GBGM, similar to the current study. They described the course of depressive symptomatology (the proximal outcome) in two groups following either Cognitive Behaviour Therapy (CBT; $N$ = 199) or 'treatment as usual' (TAU, considered to encompass the non-specific elements common to many psychological interventions, such as patient therapist relationship; $N$ = 226). Performance trajectories were related to time (measurements across 18 months) through a regression function using the continuous parameters, the intercept and slope. The intercept represented the level of depressive symptoms at pre-treatment. The mean changes in depressive symptoms over time were accounted for by the slope factor (*i.e.*, linear or quadratic slope). Their results showed that, on average, patients in both test conditions demonstrated significant improvements in depressive symptom trajectories from baseline to the 18-month follow-up, and no significant difference was found between the conditions.

West & Hastings (2011) examined a distal outcome following a training program. They used multiple conditional LGMs to assess memory intervention outcomes in older adults across a 9-week follow-up period. Both time-varying and time-invariant covariates of memory performance were analysed. Time-invariant covariates included baseline characteristics (*e.g.*, age) and were used to predict baseline memory performance (*i.e.*, the intercept was regressed on age). Time-varying covariates were also included in the model. For example, to address whether change in memory self-efficacy predicted change in memory performance across time following training, memory performance was regressed on memory self-efficacy. The researchers found that overall performance was significantly predicted by age and memory self-efficacy, and training-related gains in performance were best predicted directly by change in self-efficacy in a text recall task.

Figure 2 demonstrates a conditional LGM. It addresses a research question of whether cognitive training had a predictive effect on cognitive trajectory. The slope is regressed on the cognitive training status covariate. Cognitive training status (*i.e.*, experimental versus control) is included in the model as a fixed, dichotomous covariate to predict the slope. For simplicity this is not shown in Figure 2. There is also no distal outcome included in the model.

In sum, LGM is the simplest form of trajectory modelling. The model parameters of the intercept and slope – such as the mean, variance, covariance, linearity or non-linearity, and the outcome parameters (*i.e.*, the residual variance and measurement error) – aid in examining the characteristics of performance trajectories. Including covariates (when variances around the means are allowed) can further explain the trajectory and distal outcomes used to investigate certain consequences influenced by the growth process.

Latent growth modelling, therefore, is a useful tool when exploring a single, performance growth curve across time. Latent growth modelling also serves as a valuable statistical basis from which to highlight some of the core components of GBGM. Furthermore, LGM provides a good base from which to build more complex, conditional and group-based models, particularly for beginners to GBGM techniques. Given that it incorporates an implicit assumption that the data are from a single, homogenous population, however, LGM is not always appropriate as a final model from which to draw conclusions when population heterogeneity is suspected. Instead, GBGM techniques may be more appropriate.

## Group-based Growth Modelling

Group-based growth modelling provides a more accurate fit when heterogeneity is known or suspected within the data compared to LGM (Duncan et al., 2002; Nagin & Odgers, 2010), by relaxing the single population assumption. Group-based growth modelling treats the existence of distinct trajectories as an empirical question, addressed by exploring the developmental data. In the past, heterogeneity was addressed through specifying developmental subtypes *a priori*, based on developmental characteristics. The *a priori* approach downplays the uncertainty of classification and assumes that there are indeed distinct subtypes. Group-based growth modelling incorporates classification uncertainty into its results, thereby providing a more conservative test of potential differences across subgroups than the *a priori* approach (Connell & Frye, 2006).

**Model Parameters and Procedures in Group-based Growth Modelling**

As noted, GBGM assumes individuals are members of a finite number of latent subpopulations, or *classes*, defined by their developmental trajectories (also referred to as categorical latent class variables, or group variables) (Duncan et al., 1999; Duncan et al., 2000; Muthén, 2004; Nagin & Odgers, 2010). Heterogeneity is considered in the growth model framework through the inclusion of these classes (C; see Figure 2).

**Latent Class Growth Analysis and Growth Mixture Models**

Like LGMs, GBGMs can take different forms. The two most common methods are *latent class growth analysis* (LCGA), also known as semi-parametric group based modelling, *group-based trajectory modelling* (GBTM; Muthén & Muthén 2000; Nagin, 2005; Nagin & Land, 1993; Nagin & Odgers, 2010; Roeder, Lynch, & Nagin, 1999), or *growth mixture modelling* (GMM; Muthén, 2002; Muthén & Shedden, 1999). Statistically, a key difference between the models is the consideration of variance around the trajectories. Latent class growth analysis does not allow for variance and therefore essentially represents multiple traditional LGM trajectories

(Muthén & Muthén, 2000). Specifically, Nagin (2009) agues that that the existing variance in the overall trajectories is explained by way of the multiple latent classes. The mean growth curve for each latent class is estimated based on the initial status (intercept) and slope (linear or quadratic) for each class. Population variability is captured by differences across groups in the shape and level of their trajectories (Haviland et al., 2007; Haviland et al., 2008). In contrast, GMMs allow individual variation around the different latent growth curves (Muthén & Muthén, 1998–2010). This within-class variation therefore represents an extension of the traditional LGMs with parameters set in a mixed or multi-level random-effects model noted above (Nagin & Odgers, 2010). That is, within-class variation is represented by random effects. Whilst some researchers view LCGA and GMM as being statistical variations of the same type of analysis (*e.g.,* Muthén, 2004), others view them as addressing different research needs (*cf.* Nagin & Odgers, 2010). This is further discussed in the discussion of limitations and controversies of GBGM below. Regardless, it is often common for researchers to begin with LCGA, a more simplified model, before attempting GMMs (Connell & Frye, 2006; Jung & Wickrama, 2008; Muthén, 2004; Muthén, 2005). Both models can be produced by current software, including MPlus (Muthén, & Muthén, 1998–2010). Both were carried out in the present empirical study.

Furthermore, as previously mentioned when discussing LGM, separate group-based growth models for control, experimental and joint groups (*i.e.* the entire study sample) are conducted because, as previously noted, the groups may be regarded as different populations representing different growth, particularly due to the non-randomisation of the sample. For example, with non-randomisation of the groups, there may be other causal factors influencing cognitive performance trajectories. It is therefore necessary to determine the number of distinct trajectory classes for each population (Jung & Wickrama, 2008).

## Determining the Optimal Model

Determining the optimal number of latent classes, and therefore the model of best fit, is a key decision in GBM. Whilst there is much discussion and debate around this issue, both empirical and theoretical factors should be considered (Nagin & Odgers, 2010).[30] The empirical considerations should focus on a number of model fit indices, including the Bayesian Information Criterion (BIC; Schwartz, 1978) and the sample-size adjusted BIC (ABIC; Sclove, 1987). Models are produced in which additional classes are specified and comparisons of BIC and ABIC values are made between these consecutive models (Muthén & Muthén, 2000). This procedure continues until the lowest BIC and ABIC values are found, indicating a more optimal model. That is, these values help indicate that the model with '*K*' classes[31] is superior to the model with '*K–1*' classes. In some cases, models BICs with lower than zero values are created. Here, again, the model with the lowest value (*i.e.*, most negative value) is empirically preferred (McCutcheon & Mills, 2007). The aim is to determine the model with the lowest BIC value, although inclusion of too many classes may cause failure of model convergence and/or render the model uninterpretable (Petras & Masyn, 2010). In that case, models with a lowest BIC values are not selected.

More recently, consideration of the Vuong-Lo-Mendell-Rubin likelihood ratio test (LMR) the Lo-Mendell-Rubin Adjusted likelihood ratio test (Adjusted LRT; Lo, Mendell, & Rubin, 2001) is recommended (Muthén & Muthén, 2000). A similar

[30] This will be further discussed in the limitations and current debate section of this chapter. Class enumeration has also been discussed extensively in prior work (*cf.* McLachlan & Peel, 2004; Muthén, 2004; Nagin, 2005; Nylund, Asparouhov, & Muthén, 2007).
31 K = Number of classes (Duncan & Duncan., 2009).

technique is the the bootstrapped LRT (BLRT), suggested by McLachlan and Peel (2000). A significant *p* value for the LRT, the Adjusted LRT and bootstrapped LRT is sought (Feldman, Maysn & Conger, 2009). For each of these tests, a non-significant *p* value usually indicates no improvement of the model with *K* classes from the model with *K–1* classes. The significance level is often set at $p < 0.05$, two-tailed.

In addition, entropy can guide optimal model determination (Greenbaum, Del Boca, Darkes, & Muthén, 2005; Qureshi & Fang, 2010; Wang, 2007). Whilst not an indication of model fit, entropy indicates the degree of separation of the classes and the certainty of participants' allocation to the classes. That is, it shows the degree to which each class represents a homogenous trajectory that is different from the other classes (Petras & Masyn, 2010). Entropy is based on a summary of the posterior probabilities. Posterior probabilities provide an indication of the likelihood of a specific individual falling into a specific class in the model for the whole sample. Entropy scores can range from 0 to 1, with values closer to 1 indicating less classification errors (Nagin & Odgers, 2010). Consideration of entropy is not always endorsed, particularly given that there are no consistent cut-off criteria for deciding whether the value is reasonably high. Some authors, however, follow guidelines suggesting that a value ≥ 0.8 is acceptable (Wang, 2007).

Finally, average latent class probabilities can be used to assess the adequacy of individuals' fit to each class. Models are also often considered reliable if they include classes with more than 1% of the total participant count (Jung & Wickrama, 2008).

Despite a number of empirical considerations when determining the optimal GBGM model, selection based only on formal statistical criteria may in fact lead to an inferior choice (Nagin & Odgers, 2010). Theoretical consideration of the classes is also suggested, whereby the researcher should keep in mind the fundamental questions being addressed and the available data (Jung & Wickrama, 2008; Nagin & Odgers, 2010). It is also preferable to have *a priori* knowledge concerning the number, shape and size of trajectories (Andruff, Carraro, Thompson, Gaudreau, & Louve, 2009; Bauer, 2005). Thus, the formal statistical criteria applied to model selection helps guide subjective judgment of models of best fit.

### Exploring Inter-individual Differences: Two-stage and Generalised Approaches to Group-based Growth Modelling

When conducting GBGM, a two-stage or generalised (one-stage) approach to GBGM can be adopted (Huang et al., 2010; Muthén, 2002; Nagin, 1999).[32] Both methods enable inter-individual differences in trajectories to be explored and explained, by including covariates such as the effect of an intervention or baseline characteristics on an outcome measure (*e.g.*, Hix-Small et al., 2004; Muthén et al., 2002; Rodriguez et al., 2005; Sterba et al., 2007; Stulz et al., 2010). Some researchers, particularly in the early procedural stages of GBGM application, opt to use both the generalised and two-step methods and compare the results (Huang et al., 2010; Jung & Wickrama, 2008).

The *two-stage approach* is considered to be the conventional approach. Firstly, the optimal number of classes is identified (using either LCGA or GMM). These classes are compared based on initial performance levels or subject-specific background characteristics using conventional statistics, such as individual-sample *t*-tests, $\chi^2$, ANOVA, or multinomial logistic regression analyses (Huang et al., 2010; Langbaum

[32] A three stage approach has also been developed, although discussion of this is outside the scope of the current review.

et al., 2009; Nagin, 1999; Nagin et al., 2003; Stulz et al., 2010; Uher et al., 2010). This approach is often used with LCGA analyses when GMMs produce non-convergence, as well as small or negative variances, as is common with GMM (Qureshi & Fang, 2011).

Langbaum and colleagues (2009) carried out this two-stage approach through use of multinomial logit modelling (univariate and multiple polytomous logistic regressions). They examined memory-trained individuals from the ACTIVE study to determine if the baseline demographic and cognitive factors were predictive of distinct patterns of responsiveness to training. Their study therefore identified both heterogeneity in the sample (with three distinct performance trajectories), and found that baseline memory and speed of processing, age and education were all predictive of these distinct response patterns.

Alternatively a *generalised (one-step) approach* can be created, through the direct inclusion of covariates into a GMM. This approach is hereafter referred to as *Generalised Growth Mixture Modelling* (Muthén et al., 1998; Muthén & Muthén, 1998–2010; Muthén & Shedden, 1999). The technique is also labelled generalised group-based growth modelling (GGBGM) or conditional latent trajectory modelling (Acock, 2005; Jung & Wickrama, 2008; Muthén et al., 2002). This integrated method allows simultaneous examination of covariates' impact on longitudinal trajectories, rather than considering covariates as outcomes in *post-hoc* comparisons, as is carried out in the two-stage model (Connell & Frye, 2006; Muthén et al., 2002; Muthén et al., 2004). The integrated aspect of the model is one of the most pertinent features of GGMM (Muthén et al., 2002; Muthén et al., 2004).The GGMM (one-step) approach can also be used to predict an outcome from growth, a distal outcome or outcome indicators (Duncan et al., 2002; Jung & Wickrama, 2008; Muthén 2004). That is, GGMM can be used to examine the consequences of a set of variables (separate from the variable used to form the trajectories themselves) that are influenced by the growth process (Huang et al., 2010; Muthén et al., 2002; Muthén, 2004). Like the addition of covariates, distal outcomes can also have implications on individuals' class membership.

Procedurally, like the two-stage approach, GMM is first fitted to determine the number and shape of distinct trajectory groups (Huang et al., 2010; Jung & Wickrama, 2008; Muthén, 2004). A generalised model is then created when the covariates, growth factors (I and S) and/or the class (C) are regressed onto the covariate(s) (Jung & Wickrama, 2008; Muthén, 2001; Muthén & Muthén, 2000). Covariates can be time-invariant (*e.g.*, intervention status or baseline characteristics) or time-varying (Acock, 2005; Chen & Cohen, 2006). The empirical question dictates whether the growth parameters and/or the class are regressed on the covariate. That is, the researcher determines whether the effect will be demonstrated on the intercept (baseline level), the slope or the class membership itself. If, for example, intervention status is being evaluated, it can be entered into the model as a dichotomous predictor (*i.e.*, 'dummy coded' as the intervention group versus controls), thereby predicting whether intervention effects can explain variability across each growth trajectory. The inter-individual, temporal information with regard to (any) changes experienced following the intervention can then be provided to participants using GGMM results (Gueorguieva et al., 2007; Kreuter & Muthén, 2007). Furthermore, such information may allow for more cost-effective individual selection for training programs to maximise the cognitive gains longer-term (Stulz et al., 2010).

Muthén and colleagues (2002) used a GGMM to assess a randomised preventative intervention in Baltimore public schools, aimed at reducing aggressive behaviours in the classroom. Specifically, the analysis examined intervention effects on the slope of the developmental trajectory, that is, the change in aggressive behaviour across latent classes. Stulz and colleagues (2010) generated GGMMs to examine predictors of latent classes. Their study examined differential effects of psychosocial interventions on specific subpopulations of individuals diagnosed with cocaine dependence. They also examined baseline characteristics and determined that patient baseline characteristics, including environmental/social problems, discriminated between classes.

To create a GGMM examining an outcome from growth, a distal outcome measure is added to the model via its regression (*e.g.*, logistical regression or multinomial logistical regression) onto class membership (Jung & Wickrama, 2008; Muthén et al., 2002). Therefore, the probability of a distal outcome varies across the classes. The distal outcome measure can be continuous (Y) or dichotomous (U). Procedurally, it is recommended that the model first be fitted without the distal outcome, and compared with the model that incorporates the auxiliary information from the distal outcome (Huang et al., 2010).

Whilst this method is not highlighted in the current study or the worked example, it has been carried out by a number of researchers (*e.g.*, Duncan et al., 2002; Muthén & Muthén, 1998–2010; Muthén et al., 2002). Muthén and colleagues (2002) applied this technique to investigate the intervention effects of the program for reducing aggressive classroom behaviour. They determined the probability of juvenile delinquency as a distal outcome after their intervention and found one of the identified classes was at a significantly higher risk of delinquency compared with another class. They also identified that some of the classes were not distinguished based on this distal outcome.

A GGMM is demonstrated in two forms in Figure 2. It shows the initial GGM created extended by adding a covariate – 'cognitive training status' (a dichotomous covariate indicating the experimental or control groups) – as a predictor of slope. That is, these analyses were used to investigate effects of cognitive training (compared with controls) on each class.[33.] Figure 2 also shows that Class (C) was regressed on the covariates age, sex, education and premorbid IQ to examine if these baseline characteristics predicted class membership. To minimise the complexity of the models, these two models were created separately. Similarly, the models were considered linear and no distal outcomes were included.

## Limitations and Controversies of Group-based Growth Modelling

There are a number of limitations and controversies associated with GBGM techniques, as noted throughout this chapter. This is often the case for new and rapidly-evolving statistical techniques. One main area of limitation relates to *selecting the optimal model*. Optimal model selection encompasses empirical issues in selecting model fit indices and determining the number of latent classes. The *conceptualisation of the classes* also differs across research groups. *Practical issues* also impede the use of GBGM techniques. GBGM requires longitudinal data, adequate processes for handling missing data, and can produce inadmissible models. More minor issues can also be encountered, such as increased difficulty in producing meaningful graphs in generalised models. These are now further highlighted.

### Optimal Model Selection

**Empirical considerations.** Consideration of empirical factors regarding optimal growth model selection comes with some ambiguity. For example, there are currently

[33] This covariate was used in the conditional LGM and was also specified in the GGMM.

no exact guidelines for model fit criteria, and there are many from which to select (*e.g.*, BIC, ABIC, entropy, *etc.*, as previously noted). In addition, indices of model fit are relative, rather than absolute (Connell & Frye, 2006) thereby leading to some uncertainty with regard to optimal model selection based purely on these factors (Nagin & Odgers, 2010). It is recommended, however, that where possible theoretical and/or past research findings assists the researcher to guide optimal model selection (Andruff et al., 2009; Bauer, 2005; Jung & Wickrama, 2008; Nagin & Odgers, 2010).

**Determining the number of classes.** Determination of the number of classes is a key decision in GBM, is also problematic and there are also no clear guiding empirical principles despite extensive coverage in prior work (Bauer & Curran, 2003; Muthén, 2004; Nagin, 2005; Nagin & Odgers, 2010; Nylund et al., 2007).[34] There are differing guidelines for including variance around the classes and/or whether the researcher allows parameters to vary across classes. Variance in the models can result in problems with class enumeration when comparing different model types, *e.g.*, GMM versus LCGA (Lubke & Muthén, 2007; Nagin & Odgers, 2010). GMM can produce fewer trajectories than LCGA (Nagin & Odgers, 2010).

In a related issue, Bauer and Curran (2003) cautioned that detection of multiple classes may be artifactual, and simply be due to skewed or non-normally distributed data. This is of particular concern given that clinical data is often non-normally distributed (Nagin & Odgers, 2010). Artifactual detection of multiple classes may be avoided in LCGA however, given that the latent classes are a nonparametric representation of the distribution of the growth factors.[35] Latent curve growth analysis

[34] For further comprehensive discussion see McLachlan & Peel, 2004; Muthén, 2004; Nagin, 2005; Nylund et al., 2007.
[35] *cf.* Feldman and colleagues, 2009, for good examples of GBGM applications with skewed diagnostic data.

is thus considered a semi-parametric model (Muthén, 2004; Nagin, 1999; Nagin, 2005). That said, LCGA cannot be extended to a generalised model as, with no variation to explain, within-class covariates must also be dropped. A two-stage approach can be taken to analyse the effect of these covariates however this potentially limits the conclusions drawn (Feldman, Masyn, & Conger, 2009; Jung & Wickrama, 2008; Muthén, 2004). This will be further discussed below.

**Model complexity: Inclusion of covariates and/or distal outcomes.** There is debate as to whether the addition of a covariate or distal outcome creates a more optimal model. Some argue that if a covariate has significant direct effects on growth factors and/or class, an unconditional model may lead to distorted results, since the outcome measures at each time point will be incorrectly related to class (Jung & Wickrama, 2008; Muthén, 2004). That is, addition of covariates leads to better models by incorporating auxiliary information into the model. This information is thought to refine the membership classification and potentially produce a more reliable and valid solution (Muthén, Jo, & Broan, 2003; Muthén, 2004; Nagin & Odgers, 2010). In contrast, inclusion of covariates may blur the distinction between classes (Nagin, 2005). That is, depending on a specific individual's given value on a covariate, they may be assigned to one trajectory class yet have a trajectory that more closely resembles the mean of an alternative class.

Evidence regarding the impact of the addition of covariates into a model is mixed. Lubke and Muthén (2007) investigated the effects of covariates on performance of mixture models in a simulation study and found that correct class membership assignment increased with increasing covariate effects. Huang and colleagues (2010), however, investigated the influence of including a covariate and/or a distal outcome on GMM. Their empirical findings showed that when comparing BIC values an unconditional model had a better model fit than the three conditional models run. Similarly, Huang and colleagues (2010) compared the influence of including both a covariate and distal outcome in a GMM study investigating patterns of heroin use and the relationship of those patterns to mortality. They found that inclusion of a distal outcome resulted in class membership differences between the unconditional and conditional models. This was partly determined by the associations of the trajectories with the covariate and the distal outcome.

**Assumption of classes *a priori*.** An assumption of the existence of distinct classes *a priori*, whilst recommended, adds even more uncertainty to model selection, beyond the ambiguity of empirical considerations (Andruff et al., 2009; Bauer, 2005; Bauer & Curran, 2003; Huang et al., 2010; Nagin & Odgers, 2010). The assumption of classes does not allow for the testing of their presence (or absence). A lack of *a priori* information may also lead to a large degree of subjectivity when choosing an optimal group-based model (Bauer & Curran, 2003; Huang et al., 2010; Nagin & Odgers, 2010). This is particularly the case if theoretical information is not available or if it is limited. However, Muthén (2003) argues that this does not invalidate the method. Indeed, such assignment rules are generally considered reasonable (Nagin & Odgers).

**Conceptualisation of classes.** The conceptualisation of classes when selecting LCGA or GMM also differs between advocates of GBGM. There are different views regarding the inclusion variance around different trajectories and what the different trajectories represent (Jung & Wickrama, 2008; Muthén, 2004; Nagin & Odgers, 2010). As noted, some researchers view LCGA and GMM as being statistical variations of the same type of analysis. GMM demonstrates the variance around the trajectories and LCGA does not. Both analyses are often conducted in the same study as model complexity is built, if model admissibility permits (Connell & Frye, 2006; Jung & Wickrama, 2008; Muthén, 2004; Muthén, 2005). In these scenarios empirical considerations are prioritised.

In contrast, deciding whether to include variance around the means is also viewed as being based on considering the population from which the data is drawn (*e.g.*, Nagin & Odgers, 2010). In contexts where the population is thought to be composed of literally distinct sub-populations GMM should be used, as was the case in the present empirical study. LCGA places less emphasis on the conceptualisation of classes as literal, distinct entities. Instead, classes in LCGA provide an approximation of what is in all likelihood a continuous population of distribution of trajectories of unknown shape. Classes are considered to only represent similarities in the data itself and provide approximations for a more complex reality. Researchers do not often take this conceptualisation, however it lessens the importance of the empirical exactness of the final model (and class enumeration). That said, future replication of LCGA findings and creation of more complex models such as GGMMs (with the inclusion of variance as well as covariates and/or distal outcomes) both serve as a substantive purpose when used in the training context. Generalised growth mixture models provide supporting consideration of classes as representing meaningful strata (Nagin & Odgers, 2010). Utilisation of GGMMs also enables the researcher to consider the data as representing reality (see Kreuter & Muthén, 2007 for further discussion) or sufficiently so to be used to guide theory development, or clinical decision-making (Gueorguieva et al., 2007; Kreuter & Muthén, 2007).

## Practical Issues

**Requirement of large longitudinal samples.** Group-based growth modelling inherently requires the use of relatively large[36] longitudinal samples with multiple assessment points that may be expensive and time-consuming to collect (Connell & Frye, 2006; D'Unger et al., 1998; Eggleston, Laub, & Sampson, 2004; Sampson, Laub, & Eggleston, 2004).

**Financial and time costs.** Financial and time costs are seen through utilisation of GBGM, given that GBGM requires complicated program setups or specialised programs (Muthén & Muthén, 1998–2010). The process of creating many models when selecting an optimal model (*e.g.*, multiple iterations and models with and without covariates and/or distal outcomes) is also time-consuming. In addition, MPlus – the oft-used statistical program for conducting latent growth models (Muthén & Muthén, 1998–2010) – does not produce graphs in conditional models and does not provide graphs of the estimated and observed values for each class. Whilst this problem can be addressed, through using different software package – such as SPSS Inc. or Microsoft Excel, and entering class assignment information back into the original dataset to form graph formation and/or further conventional analyses comparing across the classes – this is clumsy and time-consuming. It should be noted, however, that MPlus is a user-friendly program that is accompanied by exemplary technical support; user manuals are clear and comprehensive and further online

[36] The exact sample size cannot be unambiguously stated, because this depends in part on other characteristics of the research design (e.g., complexity of the growth model, and the amount of variance explained by the model). For example, growth models have successfully been fitted to samples as small as $n$ = 12 and $n$ = 22 (Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991 and Pietzak et al., 2015, respectively), although sample sizes approaching at least 100 are often preferred (Curran, Obeidat, Losardo, 2010).

assistance is available (*e.g.*, discussion boards), which makes MPlus a very useful tool overall in GBGM.

**Missing data.** Missing data is a commonly-encountered issue to consider in longitudinal research. Whilst missing data on observed variables are considered missing at random (MAR),[37] covariates with missing values substantively affect the results because that participant will be dropped from the estimation (Huang et al., 2010).

**Model complexity: non-convergence.** Furthermore, there is no guarantee of model convergence with increased model complexity (Duncan et al., 2002). Increased model complexity occurs with the iterative addition of classes, the inclusion of variance around the means, as well as the inclusion of covariates (number and type, *i.e.*, time-invariant and/or time-varying) and distal outcomes. Increased model complexity also occurs when there are many and varied time points with a large amount of missing data (Huang et al., 2010; Muthén & Muthén, 1998–2010; Nagin, 1999).

Non-admissible models are also common with increased model complexity, such as when the data are on very different scales, there are poor starting values and/or when there are small variances or negative variances (Duncan et al., 2002; Huang et al., 2010; Muthén & Muthén, 1998–2010). Thus, desired models may not always be tenable in practice. Moreover, even when convergence is achieved other problems may arise such as additional (considerable) computation time, local solutions (*i.e.*, local maxima or minima) and overall model instability (*cf.* Hipp & Bauer, 2006; Huang et al., 2010; Jung & Wickrama, 2008).

---

[37] Muthén, Jo, and Brown (2003) discuss nonignorable missing data modelling using missing data indicators.

Some steps can be taken to address non-convergence. The researcher can change default values for the number of random sets and start iterations to higher values; this feature is available in MPlus. Whilst this can be effective, even with successful convergence local and multiple solutions can still occur (Duncan et al., 2002; Jung & Wickrama, 2008; Muthén, 2004). Reducing model complexity is another option. For example, estimating the intercept and not the slope (or vice versa), or limiting these changes to a particular target class rather than across all classes (Jung & Wickrama, 2008). Conducting a model without variance around the mean is also an option; however, as discussed, within-class covariates must also be dropped, thereby requiring a two-stage approach to analyse the effect of these covariates (Feldman et al., 2009; Nagin & Odgers, 2010), potentially representing a change in the conceptualisation of the classes.

## Summary

In sum, GBGM is a relatively new statistical technique that addresses limitations of other statistical methods by taking both a variable and inter-individual (person-centred) approach. Specifically, GBGM elucidates inter-individual differences in longitudinal growth trajectories.

There are a number of model parameters and procedures in GBGM. Group-based growth models incorporate latent variables including the intercept and slope, in addition to outcome parameters such as measurement error. In this respect they are similar to LGMs. Often LGMs are created as an initial procedural step. Importantly, however, GBGMs relax the single population assumption of LGM to test for heterogeneity. This is carried out through the inclusion of another model parameter: the class. Covariates and distal outcomes can also be included to predict intercepts, slopes and/or classes, and to ascertain if certain consequences are influenced by the different growth processes (distal outcomes).

Group-based growth modelling implementation requires empirical, theoretical and procedural considerations. Empirical considerations determine the optimal model, and include assessments across models of the BIC, ABIC values and statistical significance of the LMR and LRT (as well as other indicators such as entropy). Two types of group-based growth models – LCGA or GMM – can be formed. These models are created without or with variance around the mean trajectories, respectively. Ideally *a priori* information concerning the number and the shape of trajectories is useful.

When exploring inter-individual differences a two-stage and generalised (one-stage) GBGM can be executed. A two-step approach can use conventional statistics (such as individual-samples *t*-tests, $\chi^2$, ANOVA). This approach is taken if the complex model admissible is a LCGA. Alternatively, or as an additional comparative process, a generalised approach is taken through the increasing model complexity via inclusion of covariates and/or distal outcomes into the model.

There are a number limitations and controversy associated with GBGM techniques. Optimal model selection is one key area of controversy, encompassing empirical issues in selecting model fit indices and determining the number of latent classes. The conceptualisation of the classes and practical issues – including the need for longitudinal data, adequate processes for handling missing data and model inadmissibility – also add ambiguity to final, optimal model selection. Nonetheless, steps can be taken to address some of these issues, and GGMM is useful in both developmental research and clinical settings to predict distinct longitudinal trajectories.

### Optimal Model Chosen for the Current Study: Generalised Growth Mixture Modelling

Considering the benefits and limitations of the various modelling techniques discussed above, generalised GGMM was chosen as the central statistical model for the current empirical study. Generalised growth mixture modelling was implemented following LGM and LCGA to ensure the model was best represented with inter-individual trajectories and to identify the number of classes. Given that the GGMM technique enables covariates to be included directly into the model, one of the most pertinent aspects of the approach, the data output from GGMM ultimately enables consideration of data as being ‘real’, to be used to guide theory development, and/or clinical decision-making (Gueorguieva et al., 2007; Kreuter & Muthén, 2007; Muthén et al., 2002; Muthén et al., 2004). It was therefore chosen as the optimal model in the current empirical study to best identify and predict heterogeneous longitudinal cognitive performance trajectories following training. That is, it is utilised to explore inter-individual differences in cognitive performance and predict which individuals gain from training. Application of GGMM therefore enables the further exploration of ‘Use it or lose it’ and associated theories. The study was therefore both theoretically and statistically motivated.

The following empirical chapters elucidate the use of GGMM for these purposes to examine the effects of multidomain cognitive training, using a novel intervention program – Active Cognitive Enhancement (ACE) program. As noted in the introduction chapter, the study aimed to establish that there are interindividual differences in older adults' longitudinal VM and LTVM and EF performance trajectories. Specifically, if the cognitive training status can predict these trajectories, thereby demonstrating both specific and generalised effects of training. Furthermore, the study was designed to examine if individual baseline characteristics, such as age, sex and proxies for CR (education and estimated premorbid IQ) can predict the different cognitive performance trajectories. It was hypothesised that there would be heterogeneity in cognitive performance, based on cognitive training status, across all domains explored, and thus demonstration of specific and generalised effects of training. It was also hypothesised that younger females with greater levels of education and pre-morbid IQ would demonstrate greater cognitive gains.

# Chapter 6

# Method

## Participants

### Recruitment

Participants were community-dwelling and were recruited using a number of sources across varied media:

1. Newspaper articles referring to the concept of 'Use it or lose it' (The Sunday Tasmanian).
2. Newspaper advertisements requesting volunteers (separate advertisements for experimental and control participants in The Mercury, Advocate and Examiner newspapers, and the Newcastle Herald).
3. Radio interviews on local Hobart radio (both a commercial station and though the Australian Broadcasting Commission) in which the purpose of the study was detailed.
4. Flyers posted at Department of Health and Human Services, Tasmania (DHHS).
5. Public presentations through Hobart research institutes and the DHHS (with the additional aim of obtaining referrals from health professionals).
6. Alzheimer's Australia, Tasmania (AATas) internal referrals (*e.g.*, carers of people with dementia), including both within Hobart, TAS and Hunter region, NSW.
7. Word-of-mouth.

Participants initially contacted study organisers either via email or phone to register interest in either the intervention program or to join a control group. Although the majority of volunteers were able to commence intervention soon after contact with the program organisers, some eligible experimental participants were placed on a waiting list to be included in a later training group cohort. A number of other participants who initially indicated their intention to undertake the training but were unavailable to attend the intervention program on the specific dates of the study were allocated to the control groups.

**Participant Inclusion**

A telephone screening interview was conducted to collect information about participant demographics, rationale for interest in participation and eligibility for the study based on psychological and medical history. This information was then used to determine whether the participant met the following inclusion criteria:

1. Aged 55–85 years
2. Willing and able to commit to the time requirements of the study
3. Physically healthy (including no reported central nervous system or neurological disorders, recent stroke or head trauma with loss of consciousness)
4. Cognitively and psychologically healthy (all were screened for dementia and other cognitive impairments such as memory and EF deficits)
5. Currently experiencing no significant (untreated) mental health issues
6. Able to see, hear and use hands well enough to use a computer keyboard and mouse and participate in a group intervention (the latter applied only to experimental participants)

7. Had not undertaken or were not currently active in another memory or cognitive-related intervention study
8. Able and willing to provide informed consent.

All participants had sufficient conversational English language skills to complete the intervention and/or follow instructions relating to the cognitive assessments and questionnaires. Experimental participants who were deemed eligible at this assessment were also required to indicate their intention to attend at least 8 of 10 intervention sessions.

A total of 489 older adults registered their interest in participating in the study, of which 156 participants were deemed ineligible. Nine participants were outside the designated age range, 103 could not commit to the study requirements and 44 had health issues (including 5 who did not pass either initial or subsequent screening assessments for intact cognitive function). As commonly occurs in longitudinal studies (Gravetter & Wallnau, 2004) some eligible individuals missed some assessments or dropped out of the study (*e.g.*, due to illness). Eighteen participants did not complete at least 8 of the 10 training sessions and their data were subsequently discounted. The final sample therefore comprised a total of 315 participants at baseline. Initially, a smaller scale pilot study was conducted following which additional funding was received to extend the study. Participants were recruited from both Hobart, Tasmania and the Hunter region, New South Wales (NSW), Australia.[38] The pilot study research design meant that some participants were only assessed at baseline and 3 months ($n$ = 55) and not included in the 6- and 12-month follow-ups. Missing data is further discussed in the data analysis section below.

---

[38] Inclusion criteria were identical to the previously noted criteria.

Table 2.
*Participant Baseline Characteristics for Those Allocated to the Experimental and No-contact Control Groups*[39]

| | **Experimental Group (*n* = 253)** | | **Control Group (*n* = 62)** | | *p* value |
|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | |
| Age (years) | 66.82 | 7.07 | 65.76 | 7.19 | 0.29 |
| Sex (% Female) | 80.6 | - | 79 | - | .92[a] |
| Education (years) | 13.54 | 2.84 | 14.84 | 3.50 | .002* |
| Est. Premorbid FSIQ (WTAR) | 111.53 | 6.59 | 113.15 | 6.4 | 0.08 |

*Note:* FSIQ = Full Scale Intelligence Quotient; WTAR = Wechsler Test of Adult Reading. [a] Yate's Continuity Correction $\chi^2$ Test (2-sided). * $p < .05$.

It can be seen in Table 2 that both the experimental and control groups consisted of predominantly 'younger-old' adults (APA, 2009),[40] female participants, of High Average estimated premorbid IQ (Sattler & Dumont, 2004) and were well educated, having completed secondary education. Table 2 also shows that there were no significant differences between the control and treatment groups in terms of age, sex or estimated premorbid IQ. There was, however, a significant difference in education between the groups – the control group were more highly educated than the treatment group. This could not be accounted for in the final models used, due to the increase in model complexity resulting in non-convergence.[41]

**Ethical Considerations**

This research was approved by the Human Research Ethics Committee (Tasmania) Network of the University of Tasmania (UTas; Ref No: H10127).

[39] A no-contact control group was also used in the largest study of this type to date – the ACTIVE study (Ball et al., 2002; Rebok et al., 2014; Willis et al., 2006).
[40] Older adults can be defined as 'younger-old' (ages 65-74), 'older-old' (ages 75-84), and 'oldest old' (ages 85+; APA, 2009).
[41] This issue is addressed in the limitations section of the discussion.

The main potential risk to participants was the detection of decline in a participant's cognitive function during testing. When evidence of decline was identified on the cognitive outcome measures (and consultation was sought with a neuropsychologist), participants who indicated they wished to be informed of any cognitive changes were contacted by a counsellor or psychologist from AATas. Participants were also monitored for stress or anxiety from the challenges of the testing and during the training program through clinical observation by the testers. Performance anxiety was reported by a number of participants and in all such cases the examiner offered support and gave the participant the option to stop their testing if necessary. Other appropriate action (*e.g.*, referral to a counsellor/psychologist) was also offered, although no participants took this option.

Prior to commencement of study participation, written informed consent was obtained (see Appendix 2). Participants' data was coded and deidentified. Electronic data from the study are stored in password protected computer files, while paper tests and documents are stored in a locked filing cabinet at the UTas, Sandy Bay campus for a period of 5 years subsequent to the publication of any scholarly journal articles. Once this time has lapsed, all paper tests and documents will be destroyed using a paper shredder and electronic data securely deleted.

## Materials

### The Active Cognitive Enhancement Program

The Active Cognitive Enhancement (ACE) Program is a manualised, multi-domain cognitive training intervention.[42] It is a paper-based program delivered free-of-charge to participants to increase accessibility and decrease running costs. The program aimed to enhance a range of cognitive abilities known to decline with age, including: visual and verbal memory; attention and concentration; speed of processing; and EF in cognitively-intact, community-dwelling older adults. A main focus of the intervention was on mnemonic training and practice, given that memory loss is one of the most commonly reported and distressing aspects of normal ageing (Buckner, 2004). The ACE program was based on various cognitive training approaches, particularly brain-plasticity-based interventions and previous research (Katz & Rubin, 1999; Mahncke et al., 2006a; Weil & Small, 2007; Yevchak, Loeb, & Fick, 2007). It was also based on the conceptual framework of the 'Use it or lose it' hypothesis.

The program was funded by the Australian and Tasmanian governments though the Home and Community Care (HACC) Program at AATas, an Australian Research Council (ARC) linkage grant, and the Department of Health and Human Services (DHHS).

Each ACE group was facilitated by two qualified health care professionals (*e.g.*, psychologists, social workers, nurses, counsellors or other allied health professionals). All facilitators completed facilitator training and used the ACE manual and resources at each session. Some intervention groups had additional assistants (*e.g.*, psychology

[42] Treatment outcomes were interpreted as being a product of the program as a whole (*i.e.*, a combination of multiple components of the program). Some consideration of specific effects of components is made in the Discussion, where supportive evidence was available.

and social work students gaining practicum experience). At least one regular facilitator was available to facilitate every session. The intervention was predominantly conducted in central Hobart, Tasmania but also at community centres across greater Hobart to increase socio-economic outreach. A group run in the Hunter region of New South Wales (NSW) group was conducted at Alzheimer's Australia, NSW Hunter region headquarters.

The program consisted of one 2–2.5-hour training session/week, over 10 weeks.[43] Homework was also assigned over the training period, and included practise of skills and principles introduced (such as rehearsal of mnemonics). Each group consisted of approximately 20 participants, ranging from 17 to 22 allocated per group. Participant illness and group attrition altered these numbers throughout the interventions slightly. The program content was the same each week, as shown in Table 3. Detailed descriptions of each component are provided in Appendix 1.

Table 3.
*Content of Active Cognitive Enhancement (ACE) Program Training Sessions*

| **Activity** | **Approx. duration (minutes)** |
|---|---|
| Review homework Introduction to program, Session 1 | 20 |
| Educational lecturette | 25 |
| Physical activity | 5 |
| Memory strategy | 10 |
| Word memory exercise | 20 |
| *Break* | 15 |
| Visual memory exercise | 10 |
| Arithmetic exercise | 5 |
| Relaxation/Meditation | 15 |

Minor changes in program content were made between ACE groups, based on participant feedback and facilitator observations regarding the efficacy of some

[43] Results of a meta-analysis by Valenzuela & Sachdev (2009) suggested that a discrete 'dose' of cognitive exercise in the order of 2–3 months may have long-lasting and persistent protective effects on cognition over a number of years in healthy older individuals.

program elements. Changes included updates in educational material based on ongoing research, the addition of supplementary psychoeducation regarding negative cognitions (‘self-talk’) and realistic goal setting in relation to ACE program tasks and memory performance. There were also increases in allocated completion time allowed for memory tasks. Program alterations were not believed to have significant effects on the overall training outcomes – the central cognitive domains targeted with training remained and the program was thus considered to be equivalent across groups.

**Measures of Cognitive Function**

Participants were administered a battery of standardised neuropsychological tasks and subjective outcome measures, with the *Rey Auditory Verbal Learning Test* (RAVLT; Rey, 1941, 1964) and the Groton Maze Learning Task (GMLT; CogState Ltd, 2015) used as the primary outcome measures. These tests were chosen using the following criteria:

1. Adequate psychometric properties
2. Sensitivity to measure cognitive deficits known to decline in older adulthood (*i.e.*, proportion of cognitive deficits which are correctly identified, such as dementia processes)
3. Ability to measure change over time (*e.g.*, the availability of alternative versions and robustness against practice effects)
4. Availability of age- or education-corrected norms
5. Reasonable efficiency and suitability to meet the practical demands of the study. In particular, to assess training effects on VM and executive functioning, the earliest cognitive domains to show decline, which impact daily functioning, are the most subjectively worrying aspects of ARCD, and

impact on quality of life (Boron et al., 2007a; Carey, 2007; Deary et al., 2009a; Greene & Williams, 1996; Kramer & Willis, 2002; Lawton, 1982; Mahncke et al., 2006b; Thompson & Foth, 2005).

6. Use in randomised controlled trials (RCTs) studying older adults, both healthy and cognitively impaired (*e.g.*, Langbaum et al., 2009).
7. The total number of tests selected was of an appropriate the scale for the present study

**Objective cognitive measures.**

***The Rey Auditory Verbal Learning Test.*** The RAVLT has been widely used in clinical and research contexts, being considered a valid and effective measure of verbal episodic memory and learning (de Paula et al., 2012; Gross & Rebok, 2011; Magalhães & Hamdan, 2010; Espe-Pfeifer & Wachsler-Felder, 2002). The RAVLT tests verbal episodic memory, including LTVM performance (Rey, 1941, 1964). Specifically, the RAVLT evaluates an individual's ability to encode, consolidate, store and recall verbal information (Lezak et al., 2004; Magalhães, Malloy-Diniz, & Hamdan, 2012; Strauss, Shermann, & Spreen, 2006). The RAVLT is a "supraspan" task that involves serial learning of 15 aurally presented unrelated nouns (List A) across 5 trials. There is also an interference list (List B) and immediate recall of List A. Delayed recall and recognition trials are also included (the number of words recalled after a 20 minute delay and number of words correctly identified as being present or absent in the list, respectively). The RAVLT total score has high internal reliability, with a Cronbach's alpha coefficient of approximately .90 (van den Burg & Kingma, 1999). The RAVLT is also reported to have adequate test-retest reliability (Strauss et al., 2006) and is considered a valid test, correlating moderately well with other measures of memory and learning (*e.g.*, the Wechsler Memory Scale [WMS], Revised, Logical Memory Task; Johnstone, Vieth, Johnson, & Shaw, 2000). Unlike other list learning tasks – *e.g.*, the Hopkins Verbal Learning Test, Revised (HVLT-R) which has a 12 item list – the number of words in the RAVLT allows for a higher ceiling.

Administration time for the RAVLT is 10 to 15 minutes (Strauss et al., 2006). Parallel versions – Forms A, B and C – were used for follow-up assessments. Mnemonics can be applied to aid memory recall (Gross & Rebok, 2011)

The fifth trial (Trial 5) and delayed recall tasks were selected to measure VM and LTVM performance, respectively. Trial 5 is usually the best learning trial and is commonly used for memory assessment, including indication of VM performance (Klekociuk, & Summers, 2013; Uchiyama et al., 1995; Vakil, Greenstein, & Blachstein, 2010). Trial 5 was considered to represent an outcome measure most proximal to the ACE training protocols given that a large component of the cognitive training was focused on mnemonics. Whilst mnemonics can be implemented to perform the delayed recall task (*e.g.*, Fairchild et al., 2013; McKitrick et al., 1999; O'Hara et al., 2007), long-term memory was not specifically trained in the ACE program because participants were not asked to recall items in lists over a longer duration than immediate recall. As such, LTVM performance was considered to represent generalised effects of the ACE training. Trial 5 and delayed recall scores were selected given they are amongst the most reliable outcome measures (*r* values approximately .60 to .70 respectively; Uchiyama et al., 1995). Trial 5 was is also more sensitive in detecting changes in normal participants than other tests on other measures (Tierney et al., 1994).

***Groton Maze Learning Test.*** The CogState Ltd GMLT (Schroder, Snyder, Sielski, & Mayes, 2004; Snyder, Bednar, Cromer, & Maruff, 2005a; Snyder et al., 2005b; 2008) was used to assess EF. The GMLT was part a battery of CogState Ltd computerised tasks administered (www.cogstate.com). The additional CogState Ltd tasks measured a range of cognitive functions, including psychomotor speed, working memory and visual attention (CogState Ltd, 2015; Collie, Maruff, Darby, & McStephen, 2003). The CogState Ltd test battery is specifically designed for serial assessment of cognition, with multiple alternate forms and randomised presentation of stimuli (Collie, Darby, Falleti, Silbert, & Maruff, 2002; Collie et al., 2003; Makdissi et al., 2001; Westerman, Darby, Maruff, & Collie, 2001). The battery contains tests using game-like stimuli that the creators purport are intuitive and intrinsically motivating, and ensure that assessment is culture-neutral and not limited by a participant's level of education (CogState Ltd, 2008; Westerman et al., 2001).

The GMLT consisted of a 10x10 grid of tiles presented on a screen (Figure 3). Behind the tiles, a 28-step pathway was hidden. The start tile was indicated at top left and the finish shown at the bottom right of the grid. In this task, on-screen instructions ask participants to "Find the hidden pathway". Written instructions are presented on screen to indicate the task rules. Participants were instructed to use the mouse to click on the tile at the start location and then to continue find the hidden pathway, tile-by-tile, to the end. After each move, the computer indicated whether this was correct with a green tick (*i.e.*, it was the next step in the pathway). Incorrect moves were indicated with a red cross. Incorrect moves consisted of either an incorrect tile choice as the next step in the pathway, or if the participant had broken one of the rules. For example, diagonal and backward moves were disallowed. If participants made a mistake an error tone was presented. Participants were then required to click on the last correct tile and then choose a different tile to advance forward. Immediately on completion, the same maze pattern was completed four more times, for a total of five learning trials. Twenty well-matched, alternate forms of the test were available and randomly administered. The instructions encouraged participants to work as quickly as they could and be as accurate as possible.

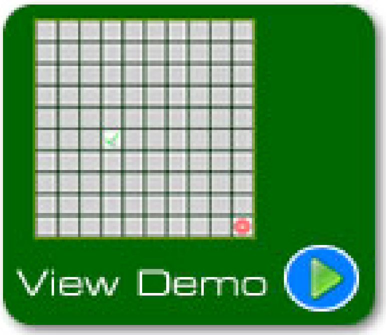


*Figure 3.* The Groton Maze Learning Test (GMLT).

Prior to the commencement of the 28-step grid, participants undertook a practice trial of the task, using a smaller grid. Instructions on the screen informed participants if they were completing a 'practice' or 'real' trial of the task. The 'practice' trial was designed by CogState Ltd to ensure that the participants understood the rules. The practice trial was shorter and easier than the 'real' test. The 'real' test was conducted immediately after each 'practice' trial. The 'practice' continued until the required number of responses was reached, or until the time period expired. The primary outcome unit of measurement was the number of errors made across the five consecutive trials. Lower scores equated to better performances.

The GMLT is considered to have good stability for the repeated assessment of neuropsychological function in older people and has been used in a number of published studies with healthy, community-dwelling older adults aged 50 years and above (*e.g.*, Ellis et al., 2009; Fredrickson et al., 2010). Thus, retest data can be considered an accurate measure of the actual level of performance (Falleti, Maruff, Collie, & Darby, 2006; Snyder et al., 2005a).

The GMLT has been validated against others measures of EF – such as the Grooved Pegboard Test and Trail Making Test (Collie et al., 2003; Maruff et al., 2009; Griffiths, McCutcheon, Silbert, & Maruff, 2006), and the Tower of Toronto and Paced Auditory Serial Addition Test (Pietrzak, Cohen, & Snyder, 2007) – with good correlations (.49–.83). In Pietrzak and colleagues' study (2007), performance on GMLT outcome measures correlated with performance on comparator measures of working memory, route selection and planning ($r$ = 0.31–0.44). These results provide support for the convergent validity of the GMLT components of EF (Pietrzak, Maruff, & Snyder, 2009).

Improvement in performance on both the RAVLT and GMLT, like other measures of cognitive performance is considered to be a manifestation of plasticity and increased CR (Brehmer et al., 2008; Noack et al., 2009; Lövdén et al., 2012; Brehmer et al., 2007). Given that these tasks were not used in the training program, any improvements were also considered to represent a level of efficiency and flexibility of cognitive processes in which participants either implicitly or explicitly were able to transfer the benefits of training (Stern, 2002) to a more generalised context.

***The Wechsler Test of Adult Reading*.** The Wechsler Test of Adult Reading (WTAR) was administered to provide estimate of premorbid IQ, which has also been shown to be a powerful proxy measure for CR (Albert & Teresi, 1999; Holdnack, 2001). It was used as a predictor in the second GGMM analysis.

The WTAR is a word-reading test where participants pronounce 50 irregularly spelled words as a baseline measure (Wechsler, 2001). The test shows excellent internal consistency (from .90 to .97 for the United States [US] standardisation sample) and is considered to be a significant predictor of intellectual functioning, with correlations to the Wechsler Adult Intelligence Test ($3^{rd}$ edition; WAIS-III) ranging from .68 to .78 for the standardisation sample of adults aged 55 and over (Holdnack, 2001). In this study a US standardisation sample and reference group, based on age, was used to convert WTAR raw scores to standard scores and predicted Full Scale Intelligence Quotient (FSIQ) scores.

***The Dementia Rating Scale.*** The Dementia Rating Scale ($2^{nd}$ edition; DRS-2; Mattis, Jurica & Leitten, 2001) was used as a screening test for undiagnosed and/or unreported cognitive deficits. It was administered at baseline and 12-month follow-up as supplementary information to ensure participants' eligibility for the study. It is considered to be well-designed for longitudinal studies (Johnson-Greene, 2004). The DRS-2 includes five subscales that provide information on specific abilities, including attention (8 items), initiation/perseverance (11 items), construction (6 items), conceptualisation (6 items) and memory (5 items). Internal consistency has also been shown to be acceptable, with a split half reliability coefficient of .90 (Johnson-Greene, 2004). Brown and Liao (1999) have shown that the DRS is valid, with most DRS subscales showing moderate to strong correlations (range = .48–.85) with common neuropsychological tests, including the WMS. In this study, the sub-scale and total raw scores were calculated and converted to age-corrected Mayo's Older Adults Normative Study (MOANS) scores. Total scores were also converted to age- and education-corrected MOANS scaled scores (AEMSS), to provide information about an individual's performance relative to others with an equivalent number of years of education. Participants with scores below 8 (indicating at least MCI) were followed up by a team psychologist and excluded from the study sample ($n$ = 8).

Other objective measures were administered to subsections of the cohort. The data from these tests were not used in the present study. Measures included:

- a buccal tissue sample taken from inside participants' cheeks by a swab by a registered nurse to investigate whether genetic factors that may modify individual risk for dementia at advanced age also played a role in the effectiveness of the ACE program in potentially improving cognitive function
- a series of tasks that measure neuromotor function. (balance and postural sway, finger tapping, hand and finger dexterity)
- answer questions on some tasks most older adults have to do in their daily life, such as taking medications, using the telephone, and money
- electroencephalography (EEG) to monitor brain activity during performance of 3 cognitive tasks.

**Subjective measures.**

***The Depression, Anxiety and Stress Scale.*** The 21-item version of the Depression, Anxiety and Stress Scale (DASS-21) was used to monitor the negative emotional states of depression, anxiety and stress (Lovibond & Lovibond, 1995). The DASS-21 is a useful measure of unique depression, anxiety and stress in older adults (Gomez, Summers, Summers, Wolf, & Summers, 2013). The three scales each contain seven items. The Depression scale assesses dysphoria, hopelessness, devaluation of life, self-deprecation, lack of interest/involvement, anhedonia and inertia. The Anxiety scale assesses autonomic arousal, skeletal muscle effects, situational anxiety and subjective experience of anxious affect. The Stress scale is sensitive to levels of chronic, non-specific arousal. A 4-point severity scale measures the extent to which each state has been experienced over the past week.

The three scales (*i.e.*, depression, anxiety and stress) in the DASS-21 have been deemed reliable with Cronbach's alpha estimated to be .88, .82 and .90, respectively (Henry & Crawford, 2004). The DASS-21 also demonstrates moderately-high to high convergent validity with the Hospital Anxiety and Depression Scale (HADS; Crawford & Henry, 2003).

Other subjective questionnaires were administered, the data from which were not used in the present study. These included: the Multifactorial Memory Questionnaire (MMQ; Troyer, 2001), a self-report meta-memory questionnaire; and the Brain Health Questionnaire (BHQ) developed specifically for the current study, which asked participants to report the number of hours spent per week in healthy lifestyle activities, *i.e.*, physical exercise, social activity, relaxation and mental activity) and monitor medication usage.[44] The Lifetime of Experiences Questionnaire (LEQ; Valenzuela & Sachdev, 2009) was also administered to a subsection of the cohort to identify factors such as education, social networks, physical activity and mentally-stimulating leisure activities. The twelve-item Control, Autonomy, Self-realisation and Pleasure (CASP-12) questionnaire was given to a sub- cohort to measure quality of life (Hyde, Wiggins, Higgs, & Blane, 2003).

[44] This information was used to monitor any medication use and/or medical conditions that may have influenced a participant's cognition and thereby warrant their exclusion from data analysis (*e.g.*, adoption of benzodiazapine use for anxiety, which has known effects on cognition).

## Procedure

### Research Design

This longitudinal study used a quasi-experimental, two-group design. Comparisons were made between individuals undertaking multi-domain cognitive training – the ACE program ($n$ = 253) – and a no-contact control group ($n$ = 62).

Training data were collated from a series of fifteen ACE program groups conducted across 2008–2011, largely within Hobart, Tasmania (and one group in the Hunter region of New South Wales, as noted as part of the pilot program).

Cognitive assessments were conducted at baseline, 3-months after baseline (immediately following the intervention program for the experimental group), and at approximately 6 and 12 months follow-up. In accordance with pilot study procedures at the initial commencement of the study, two of the 15 training groups ($n$ = 14 and 43) were assessed only at baseline and 3 months. Following obtaining informed consent (see Appendices 2–5 for participant information and consent forms). Testing was conducted at AATas Hobart headquarters, University of Tasmania (Sandy Bay Campus) and in the Alzheimer's Australia headquarters in the Hunter region, NSW. Assessment was conducted by trained allied health professionals and/or nurses and supervised by an on-team psychologist and neuropsychologist. All assessments (baseline to 12-month follow-up) involved a 2–2.5 hour testing session in rooms that were quiet and free from interruption.

During assessments, test administrators would assess overall functioning and monitor any emotional distress. The DASS-21 profiles were immediately scored and interpreted. If participants had depression scores at or above 'Severe' level (raw score

20 or above), and they indicated on their consent form if they would like to be told about psychological and cognitive deficits, their results were discussed with an ACE psychologist and participants were referred to their General Practitioner for ongoing monitoring and suggested referral to a psychologist. At the conclusion of each follow-up testing session – at 3, 6 and 12 months – participants were informed of their test results from the previous session.

Assessments were conducted at baseline and follow-up as shown in Table 4.

Table 4.
*Sequence of Test Assessments*

| Test | Details | Approx. duration (minutes) | Timepoint: Baseline | 3 months | 6 months | 12 months |
|---|---|---|---|---|---|---|
| LEQ[a] | | | ✓ | × | × | × |
| RAVLT | Trials 1–5<br>List B<br>Immediate recall List A | 10 | ✓ | ✓ | ✓ | ✓ |
| DRS-2 | | 15–30 | ✓ | × | × | ✓ |
| WTAR | | 5–10 | ✓ | × | × | × |
| RAVLT | Delayed recall/ recognition trials after 20 minutes | 5 | ✓ | ✓ | ✓ | ✓ |
| *Break* | | 5–10 | | | | |
| Subjective measures | DASS-21, MMQ, BHQ, CASP-12 | 30 | ✓ | ✓ | ✓ | ✓ |
| CogState Ltd test battery | GMLT, DET, IDN, OCL, ONB, TWOB, CPAL, GMLT-Delayed Recall | 30 | ✓ | ✓ | ✓ | ✓ |

[a] Conducted at home prior to baseline testing.
*Note*: BHQ = Brain Health Questionnaire; CASP-12 = Control, Autonomy, Self-realisation and Pleasure Scale; CPAL = Continuous Paired Associate Learning task; DASS-21 = 21-item version of the Depression, Anxiety and Stress scale; DET = Detection task; DRS-2 = Dementia Rating Scale, version 2; GMLT = Groton Maze Learning Task; IDN = Identification task; LEQ = Lifetime of Experiences Questionnaire; MMQ = Multifactorial Memory Questionnaire; OCL = One Card Learning task; ONB = One Back task; RAVLT = Rey Auditory Verbal Learning Test; TWOB = Two Back task; WTAR = Wechsler Test of Adult Reading (*cf.* CogState Ltd, 2015 for more detailed information about the CogState Ltd tasks).

## Data Analysis

### Growth Modelling: Typical Patterns of Change in Cognition Over 12 Months

Modelling of cognitive trajectories followed the recommended step-by-step approach. Model complexity was increased following the creation of a LGM (a single class model). To explore heterogeneity, GBGM was conducted, utilising LCGA and GGMM techniques. The results are presented in this order and the analysis procedures are further elaborated below. Latent growth modelling and LCGA were carried out for the control and experimental groups, and then the joint group (*i.e.*, combined data from the entire study sample) and separately for each outcome measure. As previously noted, this was conducted because the groups may be regarded as different populations representing different growth, particularly due to the non-randomisation of the sample (Jung & Wickrama, 2008; Muthén, 2005; Muthén et al., 2002). The control and experimental modelling is presented in Appendices 6–8 given that, as noted, joing group GGMM was considered optimal for the present study and was deemed the final model from which conclusions were drawn, using the joint group data.

When estimating quality and power of the GBGM techniques implemented, the sample size of the current study was considered acceptable (*e.g.*, Attix et al., 2008; Muthén, 2004; Stulz et al., 2010) and is in fact larger than some of the published literature (*e.g.*, Jones et al., 2005; West & Hastings, 2011). Monte Carlo studies of latent variable models have demonstrated that a sample size of 300 produces acceptable results (Muthén, 2004).

Generalised growth mixture modelling was conducted for VM data from Trial 5 of the RAVLT (Rey, 1941; 1964), LTVM (delayed recall of the RAVLT) and EF performance (GMLT; CogState Ltd, 2008). These measures assessed both specific and generalised training effects. Baseline characteristics – age, sex and proxies for CR (education and estimated premorbid IQ) – were included in the models to predict heterogeneity of cognitive performance trajectories.

All models were run using MPlus Version 6.12 (Muthén & Muthén, 1998–2010). Most models were conducted using MPlus defaults for parameter estimates. MPlus assumes there are residual variances and covariances for the intercept and slopes. Intercepts are fixed at zero and the intercept and slope variances are correlated. Where these settings produced inadmissible models due to non-convergence, model alterations were conducted. For example, as is common procedure, starts were increased above the initial stage random sets of starting values (from the default of 10 to 100 and the number of final stage optimisations from the default of 2 to 10) when the best log likelihood value was not replicated. This enabled determination that non-convergence was not due to local maxima. Residual variances were also fixed at zero (see Results chapters for specific modifications; Muthén & Muthén, 1998–2010; Petras & Masyn, 2010). Model results including model fit indices, baseline characteristics for each class, parameter estimates and associated figures are presented in the Results chapters to demonstrate the trajectories of these models.

**Latent Growth Modelling**

To begin the data analysis an unconditional LGM was used to demonstrate cognitive performance across 12 months (Jung & Wickrama, 2008; Laird & Ware, 1982). Latent growth modelling describes the longitudinal data by relating cognitive performance to time through a regression function, using continuous latent growth factors (the intercept and slope), as noted in Chapter 5. Change is represented by a linear slope (Jung & Wickrama, 2008; Muthén et al., 2002; Stulz et al., 2010). Quadratic modelling was not conducted due to the sample size (Acock, 2005). Non-equidistant time points (*i.e.*, initial performance, 3, 6 and 12 months) for the slope were specified accordingly. Latent growth modelling follows the simplest possibility as the null hypothesis: that a single, unconditional growth curve model can characterise the cognitive performance scores of participants across time. Model specifications were set to allow for variance (*e.g.*, Muthén & Muthén, 2002).

Conditional LGM was conducted to assess training effects on each cognitive measure for the study sample as a whole. Here, a dummy treatment/control covariate was added to the model to denote the training status for each participant (training status was coded: 1 = experimental, 2 = control group). The covariate was regressed on the slope. Here, the coefficient estimate represented an increase in the log-odds of being in the control versus the experimental class, for a one-unit increase in the covariate (Jung & Wickrama, 2008).

**Group-based latent growth models.**

***Latent class growth analysis.*** In the second stage, LCGA (Jung & Wickrama, 2008; Muthén, 2001; Nagin, 1999; Nagin & Land, 1993) was used to begin to explore heterogeneity of performance trajectories, given the indication of multiple subgroups of individual cognitive performance across time (Bissig & Lustig, 2007; Yesavage et al., 1988).

As described in Chapter 5, LCGA incorporates a categorical latent class variable into the growth model framework, allowing for the identification of distinct subpopulations of individuals (*i.e.*, latent classes), whilst variances are held equal (at zero) across classes.[45] Each class includes individuals with similar cognitive trajectories. In an iterative process, additional classes were added to the LGMs initially created to find an optimal model.

Both empirical and theoretical considerations were used as a guide to determine the LCGA model of best fit, as previously outlined in Chapter 5. The empirical considerations focused on a number of model fit indices, including the *BIC* (Schwartz, 1978), the *ABIC* (Sclove, 1987), as well as the LMR and the Adjusted LRT (Lo et al., 2001). Significance was set at $p < .05$. Class numbers were also assessed based on a number of factors. Classes were accepted if they contained above 1% of the sample (Jung & Wickrama, 2008) and if they were qualitatively distinct. The intercept was the most defining growth factor, and was therefore selected to define the classes (Fandakova et al., 2012; Muthén, online discussion forum correspondence, 2013; Stulz et al., 2010). Similarly, initial performances were distinct, further supporting their appropriate use as a label for the classes (Uher et al., 2010). This was assessed using individual sample *t*-tests for the VM and LTVM outcome measures, and one-way ANOVA for the EF measure (*e.g.*, Stulz et al., 2010). Finally, the optimal number of classes was based on consistency with similar studies in the literature (*e.g.*, Langbaum et al., 2009), as is convention (Jung & Wickrama, 2008).

Baseline characteristics – age, sex, education and estimated premorbid IQ – of the individuals in each class in the optimal model were compared using, where appropriate, individual sample *t*-tests, one-way ANOVA and $\chi^2$ analyses (*e.g.*, Fairchild et al., 2013; Huang et al., 2010). Identification of baseline characteristics of participants in each class was used to help explain the growth trajectories of cognitive performance for the identified sub-groups.

[45] Treatment status is not considered in a LCGA model.

***Generalised growth mixture modelling.*** In the third and final modelling stage, GGMM (Muthén & Muthén, 2002) was conducted. The models were considered linear and no distal outcomes were included to minimise model complexity. Additional classes were added in an iterative process, from the conditional single-class models. Models were also compared to the unconditional LCGA, given that covariates can have significant direct effects on growth factors and class, thereby leading to distorted results (Jung & Wickrama, 2008; Muthén, 2004). Specifically, two GGMM procedures were conducted, outlined below.

*Generalised growth mixture modelling assessing treatment effects.* First, GGMM was used to predict training effects on cognitive performance trajectories A dummy treatment/control covariate was used to denote the training group status for each participant (as was carried out in the conditional single class analysis; training status was coded: 1 = experimental, 2 = control group). The slope of the class trajectories was regressed on training status, assuming that intervention effects are captured in the average slopes for each class and that the intervention produced a change in within-class trajectory performance from that expected for controls (Muthén et al., 2002). In this study, a significant negative slope estimate indicated superior experimental group performance compared with controls.

*Generalised growth mixture modelling assessing predictive baseline characteristics.* Second, GGMM was used to determine predictors of class membership, thereby identifying the likely profiles of the individuals in the distinct classes. Class was regressed onto the covariates (predictors) via multinomial logistic regressions. Predictors included the following baseline characteristics: i) age; ii) sex (1 = female, 2 = male); iii) years of formal education; and iv) estimated premorbid IQ. These participant characteristics are commonly investigated in the ageing literature and have been suggested to influence cognitive performance and training effects (*e.g.*, Langbaum et al., 2009). Identifying baseline characteristic predictors of the classes further assisted in explaining the ACE program treatment effects. In this model, a significant logistic regression coefficient for each variable represented an increase in the log-odds of being in a specified class versus being in the reference subgroup. Thus, if a participant had a significant baseline characteristic, there was an increased probability of that individual belonging to the specific class compared with the allocated reference class.

**Effect Size**

After the GGMM, effect sizes were calculated, by comparing the experimental and control groups' performances from baseline to 12-month period follow up. Slope estimates for each class were also calculated (Cohen's *d*; Cohen, 1988; Muthén et al., 2002; Raudenbush & Xiao-Feng, 2001; Stulz et al., 2010). For the baseline versus 12-month results calculations, Morris and DeShon's (2002) equation 8 was applied to correct for dependence between means. Overall, effect size was calculated to indicate the magnitude of any experimental effect (Valenzuela & Sachdev, 2006a).

**Missing Data**

Table 5 demonstrates the missing data in the study within the primary outcome variables.

Table 5.
*Percentage of Missing Data for Each Outcome Measure*

| **Outcome measure** | **Missing data (%)** |
|---|---|
| VM | 23.89 |
| LTVM | 24.05 |
| EF | 27.86 |
| Training status | 0 |
| Age | 0 |
| Sex | 0 |
| Education | 0 |
| Estimated Premorbid IQ | 0.06 |

*Note:* VM and LTVM assessed using the RAVLT (Rey, 1941; 1964); EF assessed using the GMLT (CogState Ltd, 2008). VM = verbal memory; LTVM = long-term verbal memory; RAVLT = Rey Auditory Verbal Learning Test; EF = executive function; GMLT = Groton Maze Learning Task.

Parameter estimates were adjusted for missing data using a robust full information maximum likelihood (FIML) estimator, a function available within MPlus (Muthén & Muthén, 1998–2010). Full information maximum likelihood is widely accepted as a pragmatic method of handling missing data and is commonly implemented in GBGM (Arbuckle, 1996; Feldman, Nicoll, & Malenka, 1999; Muthén & Shedden, 1999; Nagin & Odgers, 2010; Schafer & Graham, 2002). Full information maximum likelihood assumes data are MAR (Donders, van der Heijden, Stijnen, & Moons, 2006; Little & Rubin, 1987), which assumes the reason for missing data is either random or allows that the "missingness" may be related to variables included in the analysis (Arbuckle, 1996; Little, 1995; Little & Rubin, 2002). Whilst some of the data may be missing completely at random (*i.e.*, missingness is unrelated to any data being modelled, present or missing), FIML yields less biased estimates than other methods of missing data, such as listwise deletion (which also reduces the power of the analysis by reducing sample size) or mean imputation (Schafer & Graham, 2002). Missing at random is considered to hold even in situations where missing values have been imposed by the researcher as a part of the study design, as in the present study (Graham, Taylor, & Cumsille, 2001). It was therefore considered that the approach to handling missing data in the present study was acceptable.

In GGMM, covariates with missing values substantively affect the results because that participant is dropped from the estimation (Huang et al., 2010). However, Table 5 shows that there was only 0.06% missing data for the covariate estimated premorbid IQ, and no missing data for treatment status, age, sex or education.

**RESULTS**

# Chapter 7

# Verbal Memory Performance

## Baseline Performance Differences Between Experimental and Control Groups

Given the non-randomisation of the study participants, baseline performances on Trial 5 of the RAVLT between the experimental and control groups were compared using an independent-samples *t*-test. There were no significant differences in scores between the experimental ($M$ = 12.40, $SD$ = 2.12) and control ($M$ = 12.87, $SD$ = 2.08) groups: $t(310) = –1.59$, $p = .11$, two-tailed. Compared to established norms, both the experimental group and the control groups' initial performances can be considered to fall within the High Average range compared with same age peers (Schmidt, 1996).

## Joint Analyses[46]

### Latent Growth Modelling of the Verbal Memory Scores for the Joint Group

**Unconditional latent growth modelling.** Table 6 shows the parameters for the classes (intercepts and slopes) for the unconditional and conditional LGM for the joint group.[47] MPlus parameter default options were used. The unconditional model had variances fixed at zero. As can be seen in Table 6, the unconditional LGM (BIC = 4039.284 and ABIC = 4020.254; Table 7) shows that the study sample as a whole

[46] As noted in the Method chapter, separate analyses were conducted for the control and experimental groups (Appendix 6).
[47] For baseline characteristics of the sample (age, sex, years of education and estimated pre-morbid IQ), please refer to Table 2 'Participant baseline characteristics' in the Methods chapter.

demonstrated a significant increase in VM performance across the 12-month follow-up period ($p < .001$).

Table 6.
*Growth Parameter Estimates for the Conditional and Unconditional Verbal Memory (VM) Latent Growth Modelling (LGM) for the Joint Group*

| | **Unconditional model** | **Conditional model** |
|---|---|---|
| | **Model estimates (*SE*; *n* = 313)** | **Model estimates (*SE*; *n* = 313)** |
| Intercept | 12.524 (0.095)** | 12.527 (0.106)** |
| Slope | 0.08 (0.014)** | –0.01 (0.020) |

*Note:* Negative slope estimate value indicates superior experimental group performance compared with controls in conditional model. SE = standard error. ** $p < .001$.

**Conditional latent growth modelling.** Table 6 also shows that the conditional model (BIC = 3861.494; ABIC = 3836.121) allowed variances to vary (*e.g.*, Muthén & Muthén, 2000) and regressed the slope parameter on training status (*i.e.*, experimental *vs.* control groups). No significant differences could be seen in VM performance trajectories across the 12-month interval between those trained in the ACE program compared with controls ($p = .60$). The VM performance for this model is identified in Figure 4. The largest point of change for the joint group was seen at the 6-month assessment point. Overall, the conditional model showed that ACE training effects on VM could not be demonstrated in a single performance trajectory of the study-population.[48]

[48] For baseline characteristics of the sample (age, sex, years of education and estimated pre-morbid IQ), please refer to Table 1 in the Methods chapter.

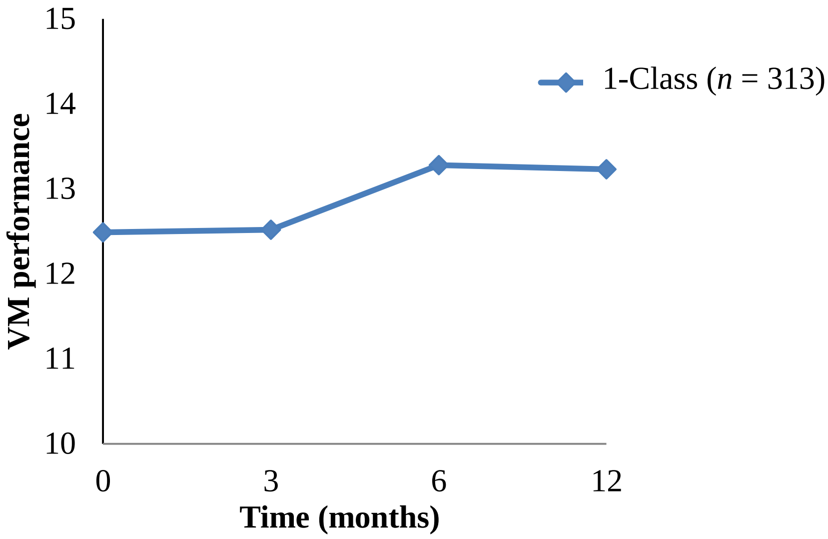


*Figure 4.* Verbal memory (VM) latent growth modelling (LGM) estimated growth trajectory across 12 months for the joint group ($n$ = 313).

**Latent Class Growth Analysis of the Verbal Memory Scores for the Joint Group**

Table 7 shows the model fit indices of the 1- to 4-class unconditional LCGA models conducted to determine the optimal model of VM performance. MPlus parameter default options were used, in addition to variances being fixed at zero. It showed support for a 3-class model, having the lowest BIC and ABIC values (BIC = 3825.299; ABIC = 3787.239) and demonstrated significant likelihood in ratio tests (LMR and adjusted LRT, $p$ = .02 and .03, respectively). The entropy value, indicating separation between the three classes, was satisfactory. Most participants were assigned to Classes 1 and 2, with Class 3 containing only 4.98% of the control cohort. While the size of Class 3 can be considered acceptable, results were interpreted with caution. Furthermore, the model demonstrated latent class probabilities of 0.908, 0.842 and 0.851 for Classes 1, 2 and 3, respectively, adding further support for the 3-class model. There was also conceptual support for the 3-class model, based on the distinct trajectories of cognitive performance across time demonstrated in past studies and the hypothesis that multiple trajectories would be identified (*e.g.*, Gross et al., 2012; Langbaum et al., 2009).

Table 7.
*Model Fit Indices from the Verbal Memory (VM) Growth Modelling for the Joint Group*

| | | | LMR | Adjusted | | Class membership (%) | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Model** | **BIC** | **ABIC** | ***p*** | **LRT *p*** | **Entropy** | **C1** | **C2** | **C3** | **C4** |
| 1-class | 4039.284 | 4020.254 | | | | 100 | | | |
| 2-class | 3892.430 | 3863.884 | .011 | .01* | 0.662 | 33.86 | 66.14 | | |
| **3-class** | **3825.299** | **3787.239** | **.02*** | **.03*** | **0.749** | **55.62** | **39.40** | **4.98** | |
| 4-class | 3833.214 | 3785.639 | .75 | .76 | 0.786 | 56.06 | 54.72 | 38.83 | 0.848 |

*Note:* Bold indicates best fit. BIC = Bayesian information criterion; ABIC = sample-size adjusted Bayesian Information Criterion; LMR = Vuong-Lo-Mendell-Rubin likelihood ratio test; Adjusted LRT = Lo-Mendell-Rubin Adjusted likelihood ratio test. * $p < .05$.

## Descriptive Variables of Classes for the Joint Group

Table 8 presents the mean (*SD*) baseline scores for the three classes for age, sex, number of years of education and estimated pre-morbid (WTAR) IQ scores for the LCGA model of the joint group. It shows a one-way ANOVA for comparison of age, years of education and pre-morbid IQ, as well as the results of planned comparisons between the classes on each of these variables in which a statistical difference was demonstrated. Pearson's $\chi^2$ was used to compare the proportions of females in each class. As can be seen in the table, there was a significant age and sex difference between the classes ($p < .001$) as well as a difference in estimated premorbid IQ ($p = .02$). Class 1 had a significantly higher mean age than Class 3 ($p = .01$), with a large effect size (Cohen's $d = 0.86$). It also had the lowest percentage of females of the three classes (40.00%). In addition, it should be noted that Class 2 was significantly older than Class 3 ($p = .001$). Class 1 had a high average IQ score, as did Classes 2 and 3, and there was a trend towards a significant difference in estimated IQ between these Classes ($p$ = .06).

Table 8.

*Descriptive Variables of the Classes from the Verbal Memory (VM) 3-Class Growth Model for the Joint Group*

| | Participant class | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Class 1 (*n* = 15)** | | **Class 2 (*n* = 125)** | | **Class 3 (*n* = 173)** | | | | | | |
| **Predictor** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***F (df)*** | ***p*** | **Pairwise comparisons** | ***p*** | **Cohen's *d*** |
| Age (years) | 70.67 | *6.36* | 68.18 | *7.43* | 65.13 | *6.59* | 9.76 (2, 310) | <.001** | 1 *vs.* 2 | .57 | 0.36 |
| | | | | | | | | | 1 *vs.* 3 | .01* | 0.86 |
| | | | | | | | | | 2 *vs.* 3 | <.001** | –0.43 |
| Female (%) | 40.00 | | 70.40 | | 91.32 | | $\chi^2$ (2, *N* = 313) = 36.74[a] | <.001*** | | | Cramer's *V* = 0.34 |
| Education (years) | 12.9 | *2.8* | 13.48 | *3.16* | 14.1 | *2.92* | 2.21 (2, 310) | .11 | | | |
| Estimated Premorbid IQ | 111.84 | *6.57* | 110.97 | *7.25* | 112.73 | *5.99* | 4.25 (2, 310) | .02* | 1 *vs.* 2 | .76 | 0.12 |
| | | | | | | | | | 1 *vs.* 3 | .09 | –0.14 |
| | | | | | | | | | 2 *vs.* 3 | .06 | 0.26 |

[a] Pearson's $\chi^2$ test; * *p* < .05, ** *p* < .001, two-tailed.

**Performance Trajectories of Classes for the Joint Group**

Table 9 presents the parameter estimates for the three classes identified. Class 3 was the largest class and had the highest intercept value; Class 2 was the second largest and had a lower intercept value; and Class 1 was the smallest class with the lowest intercept value. One-way ANOVA showed that there was a significant difference between the intercept values ($F(2, 309) = 219.729$, $p < .001$). *Post-hoc* comparisons indicated that there was a significant difference between all classes ($p < .001$). Classes 3, 2 and 1 are referred to as High, Moderate and Low VM classes, respectively. When considering the slope parameters, there was a significant positive slope for the High and Moderate VM classes while the positive slope for the Low VM was not significant (Estimate = 0.025, $SE = 0.028$, $p = .77$). Figure 5 also demonstrates the trajectories across 12 months for each class. These results suggested that VM performance for the High and Moderate VM classes increased across the 12-month follow-up period, whilst the Low VM class remained steady. No conclusions were drawn from these models. Instead they are demonstrated to represent the generalised growth mixture modelling building process, as is standard protocol (Jung & Wickrama, 2008).

Table 9.
*Growth Parameter Estimates for the Classes in the Verbal Memory (VM) 3-Class Model for the Joint Group*

| **Variable** | **Estimate** | ***SE*** | ***p*** |
|---|---|---|---|
| 1: Low VM (*n* = 15) | | | |
| Intercept factor | 8.319 | 0.634 | <.001* |
| Slope | 0.025 | 0.087 | 0.77 |
| 2: Moderate VM (*n* = 125) | | | |
| Intercept | 11.524 | 0.192 | <.001* |
| Slope | 0.063 | 0.019 | <.001* |
| 3: High VM (*n* = 173) | | | |
| Intercept | 13.617 | 0.123 | <.001* |
| Slope | 0.063 | 0.012 | <.001* |

*Note:* 1, 2, 3 indicates model class assignment in model as per Table 6. * *p* < .001, two-tailed.

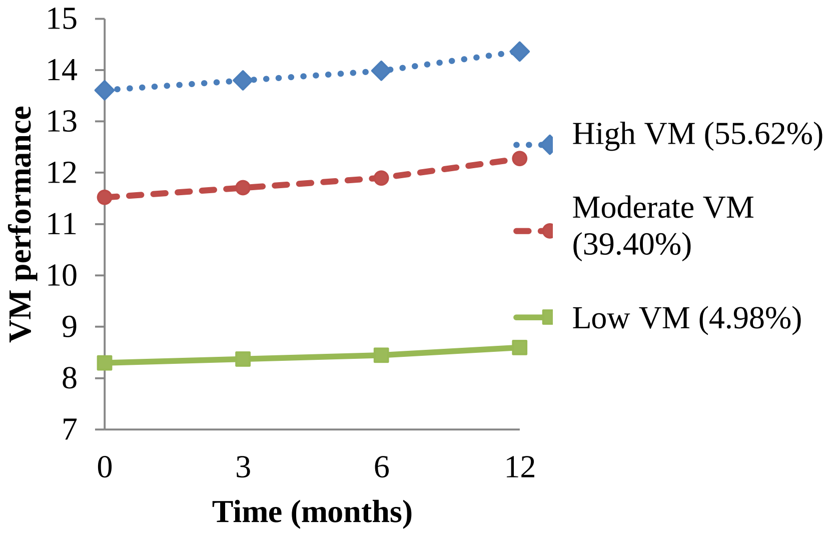


*Figure 5.* Trajectories of the verbal memory (VM) latent class growth analysis (LCGA) classes across 12 months for the joint group (*n* = 313).

**Generalised Growth Mixture Modelling with Training Status as a Predictor of Verbal Memory for the Joint Group**

Table 10 shows the various model fit indices for GGMMs that determine the optimal final model of training effects on VM performance for the joint group. MPlus parameter default options were used. As with the unconditional analyses, model assessment was conducted incrementally from 1 to 2, 3 and 4 classes and residual variances were set at zero. The model of best fit was selected from these statistical fit indices as well as conceptual considerations. Table 10 shows that a 3-class model was supported – as it was for the unconditional model – as the 3-class model had the lowest BIC and ABIC values, which improved on the unconditional model (BIC = 3841.279; ABIC = 3790.532 *vs.* LCGA BIC = 3825.299 and ABIC = 3787.239, as shown in Table 7). The 3-class GGMM model also had significant LMR ($p < .010$) and adjusted LRT values ($p < .011$). These values for the 4-class model were not significant. The entropy value indicated that separation between the three classes was satisfactory (0.749). Most participants were assigned to Classes 1 and 3. Class 2 contained only 5.11% of the total cohort and while acceptable, the results were interpreted cautiously. Furthermore, the model demonstrated latent class probabilities of 0.834, 0.882 and 0.909 for Classes 1, 2 and 3, respectively. There was also conceptual support for the 3-class model, based on the distinct trajectories of cognitive performance demonstrated in past studies (*e.g.*, Langbaum et al., 2009 Willis et al. 2006).

Table 10.
*Model Fit Indices from the Verbal Memory (VM) Growth Modelling Incorporating Training Status as a Predictor for the Joint Group*

| | | | LMR | Adjusted | | Class membership (%) | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Model** | **BIC** | **ABIC** | ***p*** | **LRT *p*** | **Entropy** | **C1** | **C2** | **C3** | **C4** |
| 1-class | 3861.494 | 3836.121 | | | | 100 | | | |
| 2-class | 3903.198 | 3868.309 | .116 | .124 | 0.664 | 32.59 | 67.41 | | |
| **3-class** | **3841.279** | **3790.532** | **.010*** | **.011**** | **0.742** | **41.21** | **5.11** | **53.67** | |
| 4-class | 3836.684 | 3776.422 | .708 | .717 | 0.69 | 34.67 | 47.30 | 5.11 | 12.922 |

*Note*: Bold indicates best fit. BIC = Bayesian information criterion; ABIC = sample-size adjusted Bayesian Information Criterion; LMR = Vuong-Lo-Mendell-Rubin likelihood ratio test; Adjusted LRT = Lo-Mendell-Rubin Adjusted likelihood ratio test; * $p < .05$, ** $p < .01$, two-tailed.

**Descriptive variables of classes in the generalised growth mixture modelling with training status as a predictor.** Table 11 presents the mean (*SD*) scores for the three classes for age, sex, number of years of education and estimated pre-morbid (WTAR) IQ scores for the GGMM incorporating training status (*i.e.*, experimental *vs.* control group) as a predictor. It shows a one-way ANOVA for comparison of age, years of education and pre-morbid IQ, and Pearson's $\chi^2$ for sex (Pallant, 2007). Table 11 also displays the results of planned comparisons between the classes on each of these variables where a statistical difference was demonstrated. As can be seen in the table, the baseline characteristics of the classes in the GGMM were very similar to those in the optimal LCGA Model. Class 2 had a significantly higher mean age than Class 3 ($p$ = .001), with a large difference (Cohen's $d$ = 2.47). Class 2 also had the lowest percentage of females of the three classes (37.50%). In addition, there was a trend towards a significant difference in education between the three classes ($p$ = .06). There was a significant difference between the classes in estimated premorbid IQ ($p$ = .03), yet all classes – including Class 1 – had, overall, a high average IQ score.

Class 2 was also significantly older than Class 3 ($p$ = .002). The difference between these groups can be considered of a large size (Cohen's $d$ = 2.47).

Table 11.

*Descriptive Variables of the Classes from the Verbal Memory (VM) Generalised Growth Mixture Modelling (GGMM) with Training Status as a Predictor for the Joint Group*

| | **Participant class** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Class 1 (*n* = 129)** | | **Class 2 (*n* = 16)** | | **Class 3 (*n* = 168)** | | | | | | |
| **Predictor** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***F (df)*** | ***p*** | **Pairwise comparisons** | ***p*** | **Cohen's *d*** |
| Age (years) | 68.09 | *7.45* | 71.38 | *6.77* | 65.03 | *6.45* | 11.18 (2, 312) | <.001** | 1 *vs*. 2 | .22 | 1.23 |
| | | | | | | | | | 1 *vs*. 3 | .001* | 1.16 |
| | | | | | | | | | 2 *vs*. 3 | .002* | 2.47 |
| Female (%) | 72.09 | | 37.50 | | 91.07 | | $\chi^2$ (2, *N* = 313) = 36.63 [a] | <.001[a]** | | | Cramer's *V* = 0.34 |
| Education (years) | 13.47 | *3.16* | 12.78 | *2.76* | 14.14 | *2.91* | 2.78 (2, 312) | .06 | | | |
| Estimated Premorbid IQ | 110.98 | *7.18* | 109.56 | *5.97* | 112.73 | *6.018* | 3.64 (2, 312) | .03* | 1 *vs*. 2 | 1.00 | 0.56 |
| | | | | | | | | | 1 *vs*. 3 | .07 | –0.68 |
| | | | | | | | | | 2 *vs*. 3 | .19 | –1.29 |

[a] Pearson's $\chi^2$ test; * $p < .01$, ** $p \leq .001$, two-tailed.

*Performance trajectories of classes in the generalised growth mixture modelling with training status as a predictor.* Table 12 shows the parameters for the classes (intercepts and slopes) for the three classes in the GGMM model assessing effects of the ACE training, comparing experimental and control groups. Class 2 was the smallest class and had the lowest intercept. Class 3 had a higher intercept value than Class 1 or 2. Thus the initial VM performance of Class 2 was lower than both Class 1 and 3. Class 1 had lower initial VM performance than Class 3. One-way ANOVA revealed that there was a significant difference in intercept values between the classes ($F(2, 312) = 214.88$, $p < .001$). *Post-hoc* comparisons using Tukey HSD indicated that there was a significant difference between all three classes ($p < .001$). Classes 1, 2 and 3 are referred to here as Moderate, Low and High VM classes, respectively. In relation to the slope parameter, the results indicate that the treatment group did not significantly predict performance trajectories for the High VM or Moderate VM classes ($p = .52$ and $.58$, respectively) compared with controls. The experimental participants in the Low VM class did, however, show a significant benefit from the intervention compared with controls ($p = .01$).

Table 12.
*Growth Parameter Estimates for the Classes in the Verbal Memory (VM) Generalised Growth Mixture Modelling (GGMM) with Training Status as a Predictor for the Joint Group*

| **Variable** | **Estimate** | ***SE*** | ***p*** |
|---|---|---|---|
| 2: Low VM (*n* = 16) | | | |
| Intercept | 8.440 | 0.702 | <.001*** |
| Slope | –0.182 | 0.074 | .014* |
| 1: Moderate VM (*n* = 129) | | | |
| Intercept | 11.598 | 0.199 | <.001*** |
| Slope | 0.018 | 0.033 | .577** |
| 3: High VM (*n* = 168) | | | |
| Intercept | 13.625 | 0.115 | <.001*** |
| Slope | –0.010 | 0.016 | .515 |

*Note:* Negative slope estimate values indicate superior performance of experimental group compared with controls; 1, 2, 3 indicates model class assignment in model as per Table 10 and Table 11. * $p < .05$, ** $p < .001$, two-tailed.

Figure 6 shows the trajectories for each of the three classes in the model. The left panel demonstrates the estimated trajectories, whilst the right panel separates the control and experimental constituents of each of the three classes and their observed trajectories across the 12-month time interval. Green lines represent the Low VM class, red lines represent the Moderate VM class and blue lines are used for the High VM class. As can be seen in the left panel, there was a significant VM performance increase in the cognitive performance slopes for of all three VM classes and the increase was greatest for all three classes at the 12-month mark. Importantly, as noted in Table 11, there was a significant difference in the slopes of the experimental and control participants in the Low VM class.

The right panel shows that the VM performance level remained consistent for both the experimental and control participants in the High and Moderate VM classes. The High

VM class was performing close to ceiling at baseline whilst the Moderate VM class was performing at an Average level at baseline (Schmidt, 1996).

Given the low *n* of the control group, an accuate measure of the magnitude of the difference between the treatment and control groups' *slopes* in the High class for VM performance could not be ascertained. Whilst there was a positive trajectory for the treatment group, ultimately training efficacy on VM was inadequately recovered.

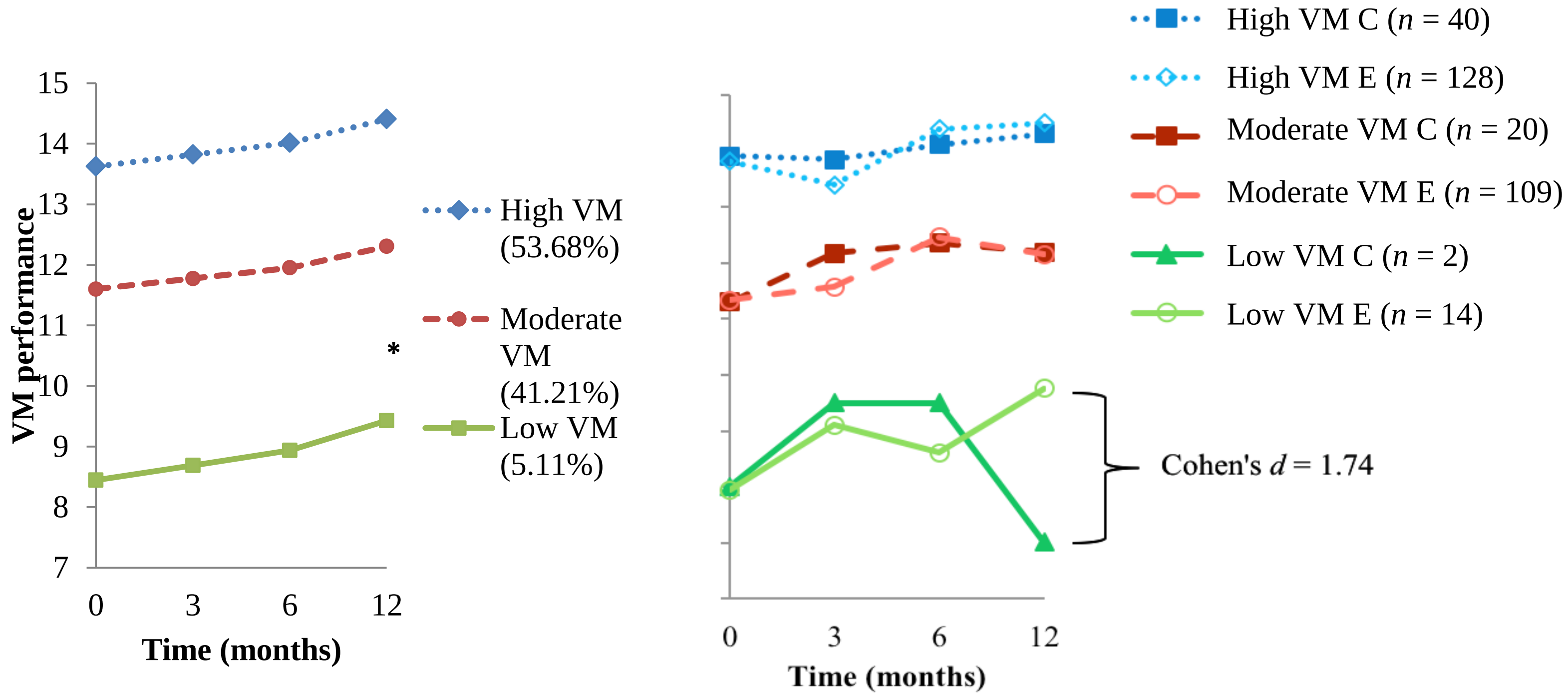


*Figure 6.* Estimated (left panel) and observed (right panel) trajectories across 12 months for the three classes of verbal memory (VM) generalised growth mixture modelling (GGMM) incorporating training status as a predictor for the joint group.
*Note:* E = experimental group; C = control group. * $p$ = .001, two-tailed.

The results of this final, generalised model for VM suggest that, participating in the ACE program training predicted statistically significant and large increases in VM performance for the experimental group in the Low VM class. Specifically, these ACE trained individuals demonstrated large improvements in VM performance across the 12-month follow-up period compared with both their initial performance and with controls. There was a large effect of training on the VM slopes for those in the experimental group of the Low VM class. However, there were only two control participants in the Low VM class to which the 14 treatment group participants were compared. The results cannot be extrapolated to other populations. The efficacy of training is therefore inadequately recovered by the GGMM.

**Generalised Growth Mixture Modelling Predicting Class Membership with Baseline Characteristics**

As previously outlined, predictors of class membership – age, sex (1 = female; 2 = male), years of formal education and estimated pre-morbid IQ – were explored in the second GGMM conducted. MPlus parameter default options were used. Table 14 shows the various model fit indices for these models, including 1- to 4-classes. As can be seen, the 3-class model was considered optimal, with the lowest BIC and ABIC values, which improved on the unconditional 3-class model (GGMM predicting class membership, BIC = 3778.002 and ABIC = 3714.569 *vs.* LCGA, BIC = 3825.299 and ABIC = 3787.239). It had significant LMR and adjusted LRT values ($p$ = .04 and .05, respectively). These values were not significant for the 4-class model. The entropy value indicated that separation between the three classes was satisfactory (0.764). Most participants were assigned to Classes 1 and 3 (55.62% and 41.21%, respectively). Whilst Class 2 contained only 6.07% of the total cohort, as previously noted, this was acceptable (Jung & Wickrama, 2008); nonetheless the results were treated with caution. Furthermore, the model demonstrated latent class probabilities of 0.906, 0.906 and 0.858 for Classes 1, 2 and 3, respectively. There was also conceptual support for the 3-class model, based on the previous GGMM of VM in the present study and the trajectories of cognitive performance demonstrated in past studies (*e.g.*, Langbaum et al., 2009; Willis et al. 2006).

Table 13.
*Model Fit Indices from the Verbal Memory (VM) Generalised Growth Mixture Modelling (GGMM) Incorporating Baseline Characteristics as Predictors of Class for the Joint Group*

| Model | BIC | ABIC | LMR *p* | Adjusted LRT *p* | Entropy | Class membership (%) C1 | C2 | C3 | C4 |
|---|---|---|---|---|---|---|---|---|---|
| 1-class[a] | 3954.288 | 3922.571 | | | | 100 | | | |
| 2-class | 3828.297 | 3787.065 | .003** | .003** | 0.761 | 29.39 | 70.61 | | |
| **3-class** | **3778.002** | **3714.569** | **.043*** | **.046*** | **0.764** | **6.07** | **41.21** | **52.72** | |
| 4-class | 3780.961 | 3695.326 | .378 | .384 | 0.863 | 14.70 | 32.59 | 46.33 | 6.39 |

*Note:* [a] Predictor effects were modelled on the slope parameter for the 1-class model (Acock, 2005); in all other models, covariate effects were exclusively modelled on class membership. Bold indicates best fit. BIC = Bayesian information criterion; ABIC = sample-size adjusted Bayesian Information Criterion; LMR = Vuong-Lo-Mendell-Rubin likelihood ratio test; Adjusted LRT = Lo-Mendell-Rubin Adjusted likelihood ratio test; * $p$ = .05, ** $p$ < .01, two-tailed.

**Descriptive variables of classes in the generalised growth mixture modelling with predictors of class membership.** Table 14 shows the mean (*SD*) baseline age, sex, years of education, and pre-morbid (WTAR) IQ scores for the three GGMM classes. It also shows the results of the appropriate statistics (one-way ANOVA for comparison of age, years of education and pre-morbid IQ, and Pearson's $\chi^2$ for sex) and the results of planned comparisons between the classes on these variables. The results indicated significant class differences for age, sex, years of education and estimated premorbid IQ ($p$ < .001). Planned comparisons of significant results indicated that for age, Class 1 was significantly older than Classes 2 and 3 ($p$ < .001). Class 1 also had the lowest percentage of females (21.2%). There was a significant and large difference in years of education between Classes 1 and 3 ($p$ = .02, Cohen's $d$ = 0.66), with Class 1 having the lowest years of education of the three classes. There was no significant difference between Class 1 and Class 2 for baseline estimated premorbid IQ ($p$ = .51). There was, however, a significantly lower estimated premorbid IQ for Class 1 compared with Class 3. This difference was of moderate size ($p$ = .02, Cohen's $d$ = –0.55).

Table 14.
*Descriptive Variables of the Classes from the Verbal Memory (VM) Generalised Growth Mixture Modelling (GGMM) with Baseline Characteristics as Predictors of Class for the Joint Group*

| | **Participant class** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Class 1 ($n$ = 16)** | | **Class 2 ($n$ = 129)** | | **Class 3 ($n$ = 168)** | | | | | | |
| **Predictor** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***F (df)*** | ***p*** | **Pairwise comparisons** | ***p*** | **Cohen's *d*** |
| Age (years) | 73.05 | *4.67* | 68.06 | *7.45* | 64.74 | *6.38* | 11.18 (2, 310) | <.001 | 1 *vs.* 2 | .008** | 0.81 |
| | | | | | | | | | 1 *vs.* 3 | <.001*** | 1.49 |
| | | | | | | | | | 2 *vs.* 3 | <.001*** | 0.48 |
| Female (%) | 21.19 | | 72.17 | | 93.97 | | $\chi^2(2, N = 313) = 67.597$ [a] | <.001[a] | | | Cramer's $V$ = 0.34 |
| Education (years) | 12.42 | *3.12* | 13.23 | *2.99* | 14.4 | *2.922* | 7.80 (2, 310) | <.001 | 1 *vs.* 2 | .51 | –0.27 |
| | | | | | | | | | 1 *vs.* 3 | .02* | –0.66 |
| | | | | | | | | | 2 *vs.* 3 | .003** | –0.40 |
| Estimated Premorbid IQ | 110.32 | *5.75* | 110.1 | *7.48* | 113.39 | *5.472* | 10.15 (2, 310) | <.001 | 1 *vs.* 2 | .51 | 0.03 |
| | | | | | | | | | 1 *vs.* 3 | .02* | –0.55 |
| | | | | | | | | | 2 *vs.* 3 | .003** | –0.50 |

[a] Pearson's $\chi^2$ test; * $p$ < .05, ** $p$ < .01, *** $p$ < .001, two-tailed.

*Performance trajectories of classes in the generalised growth mixture modelling with predictors of class membership.* Table 15 shows the parameters (intercepts and slopes) for the three classes in the GGMM incorporating baseline characteristics as predictors. Class 1, the smallest class, had the lowest intercept; Class 2 had a higher intercept value than Class 1; and Class 3 had a higher intercept value than both Class 1 and 2 (*i.e.*, Class 3 had the highest intercept value of the three classes). One-way ANOVA revealed that there was a significant difference in intercept values between the classes ($F(2, 309) = 198.294$, $p < .001$). *Post-hoc* comparisons using Tukey HSD indicated that there was a significant difference between all three classes ($p < .001$). Classes 1, 2 and 3 are referred to here as Low, Moderate and High VM classes, respectively. In relation to the slope parameters, the Low VM class showed no significant change in slope ($p = .59$). In contrast, the slopes of the Moderate and High VM classes were significant ($p < .001$). Figure 7 shows the trajectories of the slopes for each class across 12 months.

Table 15.
*Growth Parameter Estimates for the Classes in the Verbal Memory (VM) Generalised Growth Mixture Modelling (GGMM) with Baseline Characteristics as Predictors of Class for the Joint Group*

| **Variable** | **Estimate** | ***SE*** | ***p.*** |
|---|---|---|---|
| 1: Low VM (*n* = 19) | | | |
| Intercept | 8.786 | 0.687 | <.001** |
| Slope | 0.029 | 0.054 | .59 |
| 2: Moderate VM (*n* = 129) | | | |
| Intercept | 11.699 | 0.191 | <.001** |
| Slope | 0.057 | 0.019 | .003* |
| 3: High VM (*n* = 165) | | | |
| Intercept | 13.647 | 0.123 | <.001** |
| Slope | 0.063 | 0.012 | <.001** |

*Note:* 1, 2, 3 indicates model class assignment in model as per Table 14.
* $p < .01$, ** $p < .001$, two-tailed.

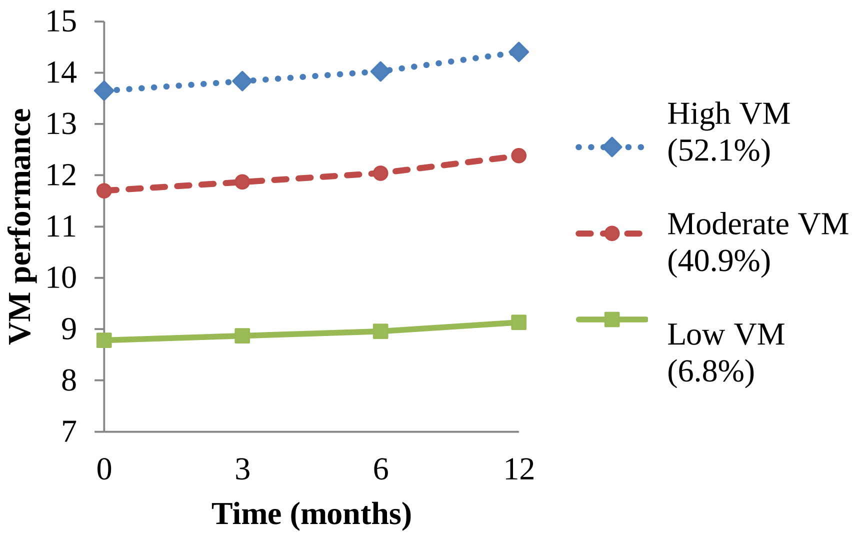


*Figure 7.* Trajectories for the 3-class verbal memory (VM) generalised growth mixture modelling (GGMM) across 12 months with baseline characteristics a predictor of class for the joint group ($n$ = 313).

Table 16 and Table 17 show the effects of the predictors (age, sex, education and estimated pre-morbid IQ) on class membership. As can be seen in Table 16, age and sex significantly discriminated between the Low and Moderate VM classes, such that individuals were more likely to be older and male if they belonged to the Low VM class ($p$ = .01 and .005, respectively). Age, sex, education and estimated pre-morbid IQ significantly discriminated between the Low and High VM classes ($p$ < .001, < .001 and $p$ = .023 and $p$ = .002, respectively). If a participant was older, male, and had a lower estimated pre-morbid IQ, there was an increased probability of that individual belonging to the Low VM class compared with the High VM class.

Table 17 shows that age, sex and estimated pre-morbid IQ significantly discriminated between the High and Moderate VM classes ($p$ = .001, $p$ < .001 and $p$ = .005, respectively). There was an increase in the probability of belonging to the Moderate VM class if participants were older (although not as old as the Low VM class), male, and had a lower estimated pre-morbid IQ compared with individuals in the High VM class. There was also a trend towards an increased probability of individuals belonging to the Moderate VM class if they had comparatively less education ($p$ = .07).

Table 16.
*Prediction of Class Membership: Class Comparisons Using the Low Verbal Memory (VM) Class as a Reference Class*

| **Variable** | **Estimate** | ***SE*** | ***p*** |
|---|---|---|---|
| Moderate VM (*n* = 129) | | | |
| Age | –0.135 | 0.055 | .01** |
| Female | –2.582 | 0.917 | .005** |
| Education | 0.209 | 0.142 | .14 |
| Estimated Premorbid IQ | 0.06 | 0.046 | .19 |
| High VM (*n* = 165) | | | |
| Age | –0.227 | 0.055 | <.001*** |
| Female | –5.054 | 0.979 | <.001** |
| Education | 0.333 | 0.146 | .023* |
| Estimated Premorbid IQ | 0.166 | 0.054 | .002** |

*Note*: Estimate represents the logistic regression coefficient (where negative values represent lower baseline characteristic values for the reference class); sex was coded 1 = female, 2 = male. * $p$ < .05, * $p \leq$ .01, *** $p$ < .001, two-tailed.

Table 17.
*Prediction of Class Membership: Class Comparison Using the High Verbal Memory (VM) Class as a Reference Class*

| **Variable** | **Estimate** | ***SE*** | ***p*** |
|---|---|---|---|
| Moderate VM (*n* = 129) | | | |
| Age | 0.092 | 0.028 | .001* |
| Female | 2.472 | 0.53 | <.001** |
| Education | 0.123 | 0.067 | .07 |
| Estimated Premorbid IQ | 0.106 | 0.038 | .005* |

*Note*: Estimate represent logistic regression coefficient (where positive values represent higher baseline characteristic values for the reference class); sex was coded 1 = female, 2 = male. * $p$ < .01, ** $p$ < .001, two-tailed. High VM *vs.* Low VM class comparisons in Table 18.

## Verbal Memory Performance Summary

In sum, the results presented here show that the conditional LGM for VM revealed no effect of training status on VM performance. The LCGA improved on the LGM, showing three distinct trajectories labelled High, Moderate and Low VM classes, based on their relative baseline performance level, indicating heterogeneity of performance (Fandakova et al., 2012; Muthén, online discussion forum correspondence, 2013; Stulz et al., 2010). The GGMM incorporating training status as a covariate showed a statistically significant and large effect of training on memory trajectory performance across 12-months for the Low VM group. However, the control sample is not considered to be of an adequate size to adequately extrapolate training specific gains to other populations. The efficacy of training on VM is therefore inadequately recovered by the GGMM.

Finally, the GGMM exploring predictors of class membership successfully showed age, sex and estimated pre-morbid IQ increased the probability that an individual was allocated to the reference Low VM class. Specifically, if a participant was older, male, and had a lower estimated pre-morbid IQ, there was an increased probability of that individual belonging to the Low VM class compared with the High VM class.

**RESULTS**

# Chapter 8

# Long-term Verbal Memory Performance

## Baseline Performance Differences Between Experimental and Control Participants

Comparisons between the raw baseline LTVM scores (the delayed recall task on the RAVLT) of the experimental and control groups were considered. There was a significant difference in scores between the experimental ($M$ = 10.31, $SD$ = 3.253) and control ($M$ = 11.58, $SD$ = 2.87) groups; $t_{309}$ = –2.82, $p$ = .005, two-tailed). Despite the significant difference between the scores, both the experimental and control groups' LTVM performances were considered to fall within the Average range compared with established norms (Schmidt, 1996).

## Joint Analyses[49]

### Latent Growth Modelling of the Long-term Verbal Memory Scores for the Joint Group

**Unconditional latent growth modelling.** Table 18 shows the parameters for the classes (intercepts and slopes) for the unconditional and conditional LGMs for the joint groups' LTVM performances across the 12-month interval. As noted in the data analysis section, MPlus parameter default options were used. The unconditional model had variances fixed at zero. As can be seen in Table 18, the unconditional

[49] As noted in the Method chapter, separate analyses were conducted for the control and experimental groups (Appendix 7).

LGM (BIC = 4742.724 and ABIC = 4765.201) showed that the study sample as a whole demonstrated a significant increase in LTVM performance across the 12-month follow-up period ($n$ = 313, Estimate = .147, $SE$ = 0.021, $p$ < .001).

Table 18.
*Growth Parameter Estimates for the Conditional and Unconditional Long-term Verbal Memory (LTVM) Latent Growth Modelling (LGM) for the Joint Group*

| | **Unconditional model** | **Conditional model** |
|---|---|---|
| | **Model estimates (*SE*; *n* = 313)** | **Model estimates (*SE*; *n* = 313)** |
| Intercept | 10.607 (0.141)* | 10.607 (0.141)* |
| Slope | 0.153 (0.02)* | 0.046 (0.03) |

*Note:* Negative slope estimate value indicates superior experimental group performance compared with controls in the conditional model. * $p$ < .001.

**Conditional latent growth modelling.** Table 18 also shows that the conditional model (BIC = 4768.550 and ABIC = 4746.348, as seen for Class 1 in Table 19), which allowed variances to vary (*e.g.*, Muthén & Muthén, 2000) and regressed the slope parameter on training status (*i.e.*, experimental *vs.* control groups). There were no significant differences in LTVM performance trajectories across the 12-month interval between those trained in the ACE program compared with controls ($p$ = .22). Overall, however, the conditional model showed that ACE training effects on LTVM could not be demonstrated in a single performance trajectory of the study-population. The LTVM performance for this model is shown in Figure 8, with highest performance evidenced at the 12-month assessment point.

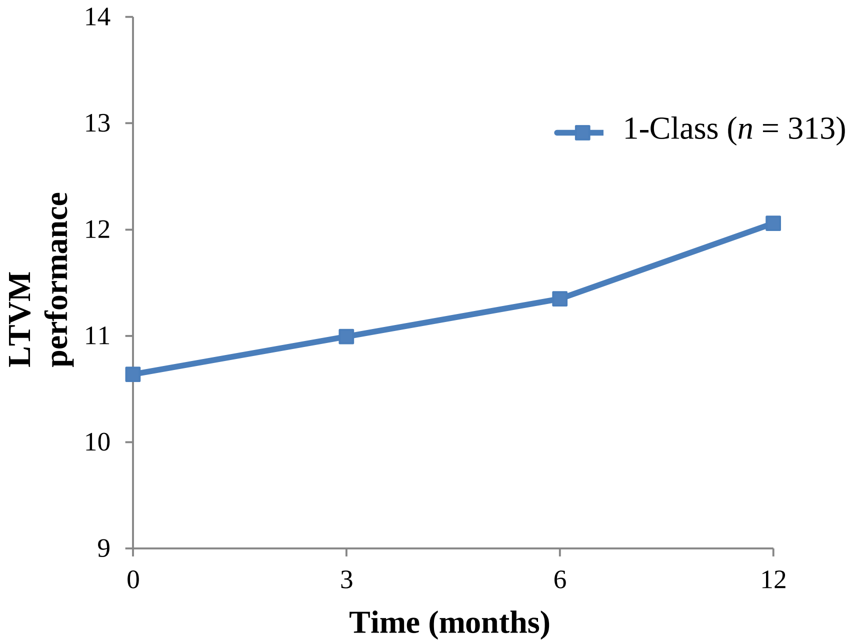


*Figure 8.* Long-term verbal memory (LTVM) latent growth modelling (LGM) estimated growth trajectory across 12 months for the joint group ($n$ = 313).

**Latent Class Growth Analysis of the Long-term Verbal Memory Scores for the Joint Group**

Table 19 shows the model fit indices of the 1- to 4-class unconditional LCGA models conducted to determine the optimal model of LTVM performance. MPlus parameter default options were used, in addition to variances being fixed at zero. Table 19 shows support for a 3-class model, which had the lowest BIC and ABIC values (BIC = 4543.728; ABIC = 4505.668) and demonstrated significant likelihood ratio tests (LMR and adjusted LRT, $p$ = .007 and $p$ = .009, respectively). The entropy value indicating separation between the three classes was satisfactory (0.720). Most participants were assigned to Classes 2 and 3. Class 1 was small – consisting of 6.71% of the control cohort – and considered acceptable, but results were interpreted with caution. Furthermore, the model demonstrated latent class probabilities of 0.873, 0.839 and 0.881 for Classes 1, 2 and 3, respectively, adding further support for the 3-class model. There was also conceptual support, based on the three distinct trajectories of cognitive performance across time demonstrated in past studies (*e.g.*, Langbaum et al., 2009).

Table 19.
*Model Fit Indices from the Long-term Verbal Memory (LTVM) Growth Modelling for the Joint Group*

| **Model** | **BIC** | **ABIC** | **LMR *p*** | **Adjusted LRT *p*** | **Entropy** | **Class membership (%)** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | **C1** | **C2** | **C3** | **C4** |
| 1-class | 4765.201 | 4746.171 | | | | 100 | | | |
| 2-class | 4601.174 | 4572.629 | .002* | .003* | 0.669 | 62.30 | 37.70 | | |
| **3-class** | **4543.728** | **4505.668** | **.007*** | **.009*** | **0.720** | **6.709** | **43.770** | **49.520** | |
| 4-class | 4544.175 | 4496.600 | .003* | .003* | 0.762 | 4.150 | 45.690 | 48.240 | 1.920 |

*Note:* Bold indicates best fit. BIC = Bayesian information criterion; ABIC = sample-size adjusted Bayesian Information Criterion; LMR = Vuong-Lo-Mendell-Rubin likelihood ratio test; Adjusted LRT = Lo-Mendell-Rubin Adjusted likelihood ratio test. * $p < .01$.

### Descriptive Variables of Classes for the Joint Group

Table 20 presents the mean (*SD*) baseline scores for the three classes for age, sex, number of years of education and estimated pre-morbid (WTAR) IQ scores for the LCGA model of the joint group. It shows a one-way ANOVA for comparison of age, years of education and pre-morbid IQ, as well as the results of planned comparisons between the classes on each of these variables in which a statistical difference was demonstrated. Pearson's $\chi^2$ was used to compare the proportions of females in each class. As can be seen in the table, there was a significant age and sex difference between the classes ($p < .001$) as well as a difference in estimated premorbid IQ ($p = .03$). Class 1 had a significantly higher mean age than Classes 2 and 3 ($p = .03$ and $< .001$, respectively). The size of these age differences was medium and large: Cohen's $d$ = 0.55 and .98 for classes 1 *vs.* 2 and 1 *vs.* 3, respectively. Class 1 also had the lowest percentage of females of the three classes (42.86%). In addition, it should be noted that Class 1 had a high average IQ score, as did Classes 2 and 3. Class 2 was also significantly older than Class 3 ($p$ = .006) and there was a trend towards a difference in estimated IQ between these classes ($p$ = .06).

Table 20.
*Descriptive Variables of the Classes from the Long-term Verbal Memory (LTVM) 3-Class Growth Model for the Joint Group*

| | **Participant class** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Class 1 (*n* = 21)** | | **Class 2 (*n* = 137)** | | **Class 3 (*n* = 153)** | | | | **Pairwise** | | |
| | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***F (df)*** | ***p*** | **comparisons** | ***p*** | **Cohen's *d*** |
| Age (years) | 71.71 | *6.97* | 67.72 | *7.49* | 65.23 | *6.28* | 10.99 *(2, 310)* | <.001** | 1 *vs.* 2 | .03* | 0.55 |
| | | | | | | | | | 1 *vs.* 3 | <.001** | 0.98 |
| | | | | | | | | | 2 *vs.* 3 | .006** | 0.34 |
| Female (%) | 42.86 | | 75.18 | | 90.32 | | $\chi^2_{(2,\, N = 313)}$ = 27.62 [a] | <.001**[a] | | | Cramer's *V* = 0.34 |
| Education (years) | 13.26 | *3.5* | 13.51 | *2.89* | 14.11 | *3.06* | 2.21 *(2, 310)* | 0.11 | | | |
| Estimated Premorbid IQ | 110.24 | *5.97* | 111.01 | *7.32* | 112.81 | *5.81* | 3.44 *(2, 310)* | .03* | 1 *vs.* 2 | 1 | –0.12 |
| | | | | | | | | | 1 *vs.* 3 | .28 | –0.44 |
| | | | | | | | | | 2 *vs.* 3 | .06 | –0.27 |

[a] Pearson's $\chi^2$ test. * $p$ < .05, ** $p$ < .001, two-tailed.

## Performance Trajectories of Classes for the Joint Group

Table 21 presents the parameter estimates for the three classes identified. Class 1 was the smallest class with the lowest intercept; Class 2 had a higher intercept and was larger; and Class 3 was the largest class with the highest intercept value. One-way ANOVA demonstrated that there was a significant difference between the intercept values ($F(2, 308) = 211.63$, $p < .001$). *Post-hoc* comparisons indicated that there was a significant difference between all classes ($p < .001$). Classes 1, 2 and 3 are referred to as Low, Moderate and High LTVM classes, respectively. When considering the slope parameters, there was a significant positive slope for the Low and Moderate LTVM classes. The positive slope for the High LTVM was not significant ($p = .70$). Figure 9 demonstrates the trajectories across 12 months for each class. These results suggest that LTVM performance for the Low and Moderate LTVM classes increased across the 12-month follow-up period, whilst the High LTVM class remained steady over this interval.

Table 21.
*Growth Parameter Estimates for the Classes in the Long-term Verbal Memory (LTVM) 3-Class Model for the Joint Group*

| **Variable** | **Estimate** | ***SE*** | ***p*** |
|---|---|---|---|
| 1: Low LTVM (*n* = 21) | | | |
| Intercept | 5.871 | 0.872 | <.001** |
| Slope | 0.148 | 0.024 | <.001** |
| 2: Moderate LTVM (*n* = 137) | | | |
| Intercept | 9.634 | 2.254 | <.001** |
| Slope | 0.088 | 0.026 | .001** |
| 3: High LTVM (*n* = 155) | | | |
| Intercept | 12.283 | 0.213 | <.001** |
| Slope | 0.044 | 0.112 | .70 |

*Note:* 1, 2, 3 indicates model class assignment in model as per Table 20. * $p < .01$, ** $p \leq .001$.

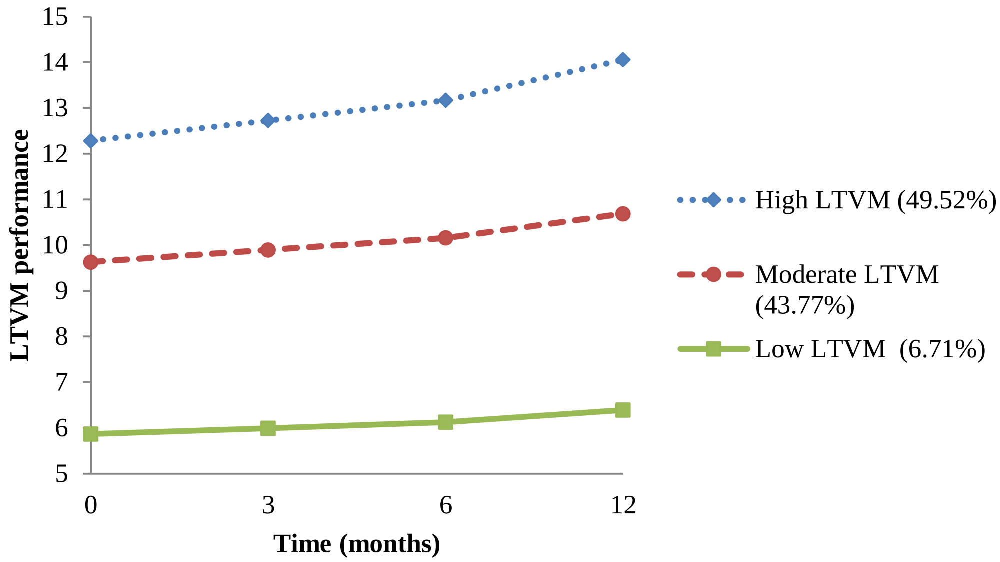


*Figure 9.* Trajectories of the long-term verbal memory (LTVM) latent class growth analyses (LCGA) classes across 12 months for the joint group ($n$ = 313).

**Generalised Growth Mixture Modelling with Treatment Status as a Predictor of Long-term Verbal Memory for the Joint Group**

Table 22 shows the various model fit indices for the growth modelling conducted to determine the optimal final model of training effects of LTVM performance for the joint group. MPlus parameter default options were used. As was previously highlighted in the unconditional analyses, model assessment was conducted incrementally from 1 to 2, 3 and 4 classes. The model of best fit was selected from these statistical fit indices as well as for conceptual considerations. Table 22 shows that a 3-class model was supported, as it was for the unconditional model. Whilst the 4-class model had the lowest BIC and ABIC values, the LMR and adjusted LRT values were not significant ($p$ = .11 and $p$ = .12, respectively). The 3-class GGMM model (BIC = 4563.150 and ABIC = 4512.404) also had significant LMR and adjusted LRT values ($p$ = .01). The entropy value indicated that separation between the three classes was good (0.72). Most participants were assigned to Classes 2 and 3, while Class 1 contained only 7.37% of the total cohort. While this can be considered acceptable, the results were interpreted cautiously. Furthermore, the model demonstrated latent class probabilities of 0.857, 0.883 and 0.834 for Classes 1, 2 and 3, respectively. There was conceptual support for the 3-class model, based on the existence of distinct trajectories of cognitive performance demonstrated in past studies (*e.g.*, Langbaum et al., 2009; Willis et al. 2006).

Table 22.
*Model Fit Indices from the Long-term Verbal Memory (LTVM) Growth Modelling Incorporating Training Status as a Predictor for the Joint Group*

| | | | **LMR** | **Adjusted** | | **Class membership (%)** | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Model** | **BIC** | **ABIC** | ***p*** | **LRT *p*** | **Entropy** | **C1** | **C2** | **C3** | **C4** |
| 1-class | 4768.550 | 4746.348 | | | | 100 | | | |
| 2-class | 4541.739 | 4500.507 | .01* | .01* | 0.58 | 29.07 | 70.93 | | |
| **3-class** | **4563.150** | **4512.404** | **.01*** | **.01*** | **0.72** | **7.35** | **49.20** | **43.45** | |
| 4-class | 4555.583 | 4495.321 | .11 | .12 | 0.69 | 45.69 | 5.11 | 34.82 | 14.38 |

*Note:* Bold indicates best fit. BIC = Bayesian information criterion; ABIC = sample-size adjusted Bayesian Information Criterion; LMR = Vuong-Lo-Mendell-Rubin likelihood ratio test; Adjusted LRT = Lo-Mendell-Rubin Adjusted likelihood ratio test. * $p = .01$, two-tailed.

**Descriptive variables of classes in the generalised growth mixture modelling with training status as a predictor.** Table 23 presents the mean (*SD*) scores for the three classes for age, sex, number of years of education and estimated pre-morbid (WTAR) IQ scores for the GGMM incorporating training status (*i.e.*, experimental *vs.* control group) as a predictor of LTVM performance across the 12-month interval. The table shows a one-way ANOVA for comparison of age, years of education and pre-morbid IQ, and Pearson's $\chi^2$ comparison for sex. It also displays the results of planned comparisons between the classes on each of these variables where a statistical difference was demonstrated. As can be seen in the table, the baselines of the classes in the GGMM were very similar to those in the LCGA model (Table 20). Class 1 had a significantly higher mean age ($M$ = 70.96 years) than Class 2 ($p < .001$), with a large difference (Cohen's $d$ = 2.23). Class 1 also had a significantly higher mean age than Class 3 ($M$ = 70.96 years, $p < .001$), with a large difference Cohen's $d$ = 1.26. Class 1 had the lowest percentage of females of the three classes (47.83% female). There was a trend toward a difference in estimated premorbid IQ between these classes ($p$ = .06), yet all classes, including Class 3, had a high average IQ score.

Table 23.

*Descriptive Variables of the Classes from the Long-term Verbal Memory (LTVM) Generalised Growth Mixture Modelling (GGMM) with Training Status as a Predictor for the Joint Group*

| | **Participant class** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Class 1 (*n* = 23)** | | **Class 2 (*n* = 154)** | | **Class 3 (*n* = 136)** | | | | | | |
| **Predictor** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***F (df)*** | ***p*** | **Pairwise comparisons** | ***p*** | **Cohen's *d*** |
| Age (years) | 70.96 | *7.16* | 65.12 | *6.52* | 67.57 | *7.34* | 9.40 *(2, 312)* | <.001** | 1 *vs.* 2 | .001* | –2.23 |
| | | | | | | | | | 1 *vs.* 3 | .09 | 1.26 |
| | | | | | | | | | 2 *vs.* 3 | .01* | –0.93 |
| Female (%) | 47.83 | | 90.26 | | 75.00 | | $\chi^2_{(2, N = 312)}$ = 27.62 | <.001**a | | | Cramer's *V* = 0.30 |
| Education (years) | 13.13 | *3.43* | 14.18 | *3.07* | 13.66 | *2.92* | 1.82 *(2, 312)* | 0.16 | | | |
| Estimated Premorbid IQ | 110.65 | *5.87* | 112.74 | *5.84* | 111.04 | *7.34* | 2.87 *(2, 312)* | 0.06 | | | |

[a] Pearson's $\chi^2$ test. * *p* = .01, ** *p* = .001, two-tailed.

*Performance trajectories of classes in the generalised growth mixture modelling with training status as a predictor.* Table 24 shows the parameters for the classes (intercepts and slopes) in the GGMM model assessing effects of the ACE training, comparing experimental and control groups. Class 1 was the smallest class and had the lowest intercept and Class 2 had the highest intercept. Thus the order of classes from lowest to highest initial LTVM performance was Class 1, 3 and 2, respectively. One-way ANOVA revealed that there was a significant difference in intercept values between the classes ($F(2, 312) = 202.86$, $p < .001$). *Post-hoc* comparisons using Tukey HSD indicated that, like the 3-class LCGA model, there was a significant difference between all three classes ($p < .001$). Classes 1, 2 and 3 are referred to here as Low, High and Moderate LTVM classes, respectively. In relation to the slope parameter, the results indicate that membership of the treatment group trended towards significantly predicting performance trajectories for the Low LTVM ($p = .067$). The experimental participants in the High and Moderate LTVM classes did not profit from training ($p = .97$ and $p = .70$, respectively).

Table 24.
*Growth Parameter Estimates for the Classes in the Long-term Verbal Memory (LTVM) Generalised Growth Mixture Modelling (GGMM) with Training Status as a Predictor for the Joint Group*

| **Variable** | **Estimate** | ***SE*** | ***p*** |
|---|---|---|---|
| 1: Low LTVM (*n* = 23) | | | |
| Intercept | 5.959 | 0.853 | <.001* |
| Slope | –0.134 | 0.071 | .067 |
| 3: Moderate LTVM (*n* = 136) | | | |
| Intercept | 9.680 | 0.274 | <.001* |
| Slope | –0.021 | 0.055 | .70 |
| 2: High LTVM (*n* = 154) | | | |
| Intercept | 12.251 | 0.260 | <.001* |
| Slope | –0.001 | 0.028 | .97 |

*Note:* Negative slope estimate values indicate superior performance of experimental group compared with controls; 1, 2, 3 indicates model class assignment in model as per Table 24. * $p < .001$, two-tailed.

Figure 10 shows the trajectories for each of the three classes in the model. The left panel demonstrates the estimated trajectories, whilst the right panel separates the control and experimental constituents of each of the three classes and their observed trajectories across the 12-month time interval. Green lines represent the Low LTVM class, red lines represent the Moderate LTVM class and blue lines are used for the High LTVM class. As can be seen in the left panel, the LTVM performance increased across the 12-month interval for all three LTVM classes and was at its greatest for all three classes at the 12-month follow-up. There was a trend towards a significant difference in the *slope estimates* of the experimental and control participants. There was large effect size (Cohen's *d* = 1.48). Thus, magnitude of the differences in the slopes between the control and experimental groups was large (Valenzuela & Sachdev, 2006a). There was also a small effect of training on the slope for the experimental group in the Moderate LTVM class (Cohen's *d* = 0.20) and an inconsequential effect of training in the High LTVM class (Cohen's *d* = 0.01).

The right panel shows that the LTVM performance level of both the experimental and control participants of the High and Moderate LTVM classes showed some improvement. However, when the two groups' performance changes from baseline to 12-months were compared, they were not significantly different. The High and Moderate LTVM classes performed at an Average level compared to norms at baseline.

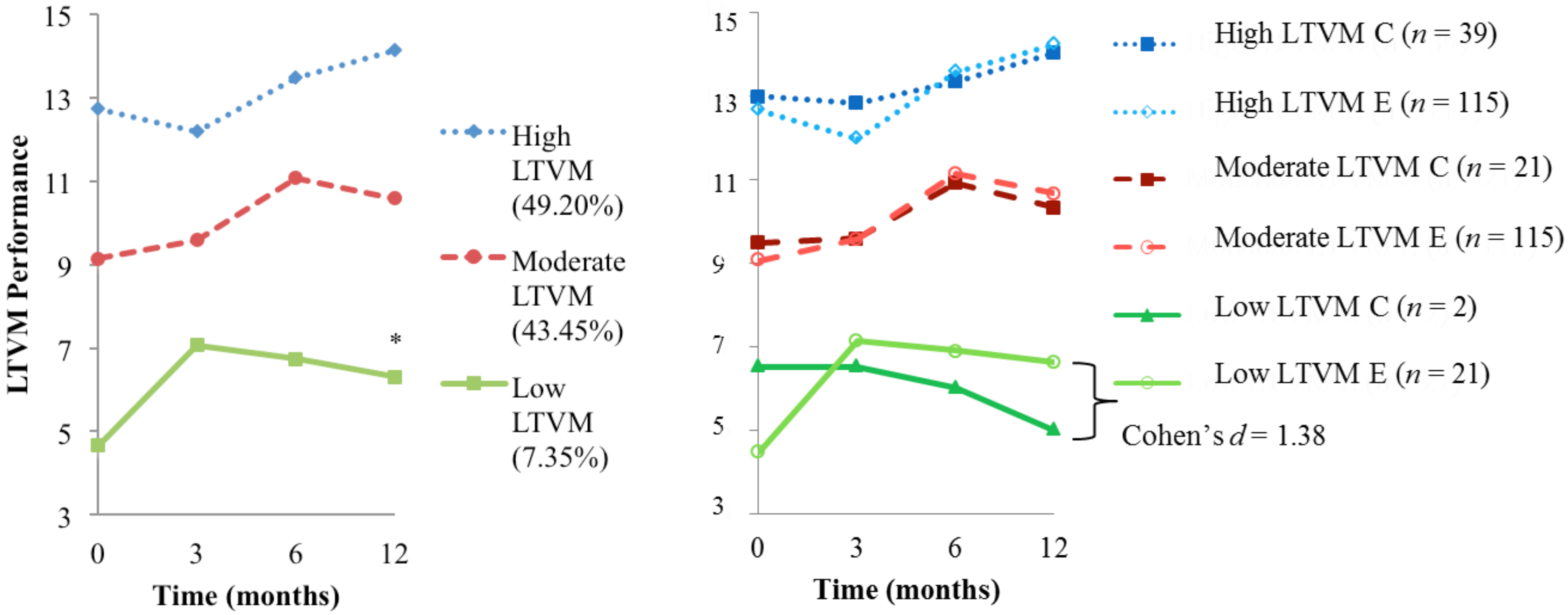


*Figure 10.* Estimated (left panel) and observed (right panel) trajectories across 12 months for the three classes of the long-term verbal memory (LTVM) generalised growth mixture modelling (GGMM) incorporating training status as a predictor for the joint group ($n$ = 313).
*Note:* E = experimental group; C = control group. * $p$ = .001, two-tailed, Cohen's $d$ = 1.48 (for slope comparisons; left panel).

Both Figure 10 and Table 25 (below) show the effect sizes of LTVM performance between the initial observed performance and 12-month follow-up for the experimental and control participants within each GGMM class. This demonstrates the magnitude of training effect across time between controls and experimental groups (Valenzuela & Sachdev, 2006a). There were larger effects on LTVM performance for experimental participants of all classes, compared with controls. The difference in the effect sizes for the treatment and control groups for the Low LTVM class was large, given the experimental group increased and the controls' performance decreased (Cohen's *d* = 1.38). There was a small effect of training for the Moderate LTVM class and a moderate effect for the High LTVM class (Cohen's *d* = 0.31 and 0.46, respectively).

Table 25.
*Effect Sizes for the Baseline to 12-month Comparisons for Experimental and Control Groups in the 3-Class Long-term Verbal Memory (LTVM) Generalised Growth Mixture Modelling (GGMM) Incorporating Training Status as a Predictor*

| | **Group** | | |
|---|---|---|---|
| **Predictor** | **Experimental** | **Control** | **Effect size difference** |
| Low LTVM | 0.67 | –0.71 | 1.38 |
| Moderate LTVM | 0.66 | 0.35 | 0.31 |
| High LTVM | 0.91 | 0.45 | 0.46 |

*Note:* Effect size, Cohen's *d*, using Morris and DeShon's (2002) equation 8 to correct for dependence between means.

The results of this final, generalised model for LTVM suggest that, participating in the ACE program training only predicted a trend towards increases in LTVM performance for the experimental group compared to controls in the Low VM class, Whilst ACE trained individuals demonstrated a large magnitude of improvement in LTVM performance across the 12-month follow-up period compared with both their initial performance and with controls, overall there were only two control participants in the Low VM class to which the treatment group participants were compared. The trend towards an improvement in LTVM from training therefore cannot be extrapolated to other populations. The efficacy of training on LTVM is therefore inadequately recovered by the GGMM.

**Generalised Growth Mixture Modelling Predicting Class Membership with Baseline Characteristics**

As previously outlined, four predictors of class membership – age, sex (1 = female; 2 = male), years of formal education and estimated pre-morbid IQ – were explored in the second GGMM used in this study. MPlus parameter default options were used. Table 26 shows the various model fit indices for these models, including 1- to 4-classes. As can be seen, the 3-class model was considered optimal, with the lowest BIC and ABIC values that improved on the unconditional 3-class model (GGMM predicting class membership: BIC = 4518.920 and ABIC = 4455.486 as seen in Table 34 *vs.* LCGA: BIC = 4543.728 and ABIC = 4505.668 as seen in Table 19). It had significant LMR and adjusted LRT values ($p$ = .006 and .007, respectively). These values were not significant for the 4-class model. The entropy value indicated that separation between the three classes was good (0.76). Most participants were assigned to Classes 1 and 3 (46.01% and 46.65%, respectively). Whilst Class 2 contained only 7.35% of the total cohort, as previously noted, this was acceptable yet results were treated with caution. Furthermore, the model demonstrated latent class probabilities of 0.903, 0.884 and 0.870, for Classes 1, 2 and 3, respectively. There was also conceptual support for the 3-class model, based on the previous GGMM of LTVM in the present study and three distinct trajectories of cognitive performance demonstrated in past studies (*e.g.*, Langbaum et al., 2009; Willis et al. 2006).

Table 26.
*Model Fit Indices from the Long-term Verbal Memory (LTVM) Generalised Growth Mixture Modelling (GGMM) Incorporating Training Status as a Predictor of Class for the Joint Group*

| | | | **LMR** | **Adjusted** | | **Class membership (%)** | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Model** | ***BIC*** | **ABIC** | ***p*** | **LRT *p*** | **Entropy** | **C1** | **C2** | **C3** | **C4** |
| 1-class | 10909.567 | 10865.165 | | | | 100 | | | |
| 2-class | 4569.533 | 4528.301 | .001** | .001** | 0.69 | 59.43 | 40.58 | | |
| **3-class** | **4518.920** | **4455.486** | **.006*** | **.007*** | **0.76** | **46.01** | **7.35** | **46.65** | |
| 4-class | 4539.400 | 4463.279 | .12 | .12 | 0.77 | 45.69 | 4.79 | 46.65 | 2.88 |

*Note:* Predictor effects were modelled on the slope parameter for the 1-class model (Acock, 2005); in all other models, predictor effects were exclusively modelled on class membership. Bold indicates best fit. BIC = Bayesian information criterion; ABIC = sample-size adjusted Bayesian Information Criterion; LMR = Vuong-Lo-Mendell-Rubin likelihood ratio test; Adjusted LRT = Lo-Mendell-Rubin Adjusted likelihood ratio test. * $p < .01$, ** $p = .001$, two-tailed.

**Descriptive variables of classes in the generalised growth mixture modelling with predictors of class membership.** Table 27 shows the mean (*SD*) baseline age, sex, number of years of education and pre-morbid (WTAR) IQ scores for the three GGMM classes. It also shows the results of the appropriate statistics (one-way ANOVA for comparison of age, years of education and pre-morbid IQ, and Pearson's $\chi^2$ for sex) and the results of planned comparisons between the classes on these variables. The results indicated significant class differences for age, sex and estimated premorbid IQ ($p < .001$), as well as years of education ($p = .005$). *Post-hoc* comparisons of significant results indicated that Class 2 participants were significantly older than those in Classes 1 and 3. These differences were of a moderate and large size, respectively (Class 1 *vs.* 2, $p < .001$, Cohen's $d = -1.26$; Class 2 *vs.* 3, $p = .006$, Cohen's $d = 0.66$). Class 2 had the lowest percentage of females (34.82%). There was a significant and moderate difference in years of education between Classes 1 and 2 ($p = .05$; Cohen's $d = -0.49$), as well as a significant and small difference in years of education between Classes 1 and 3 ($p = .01$, Cohen's $d = 0.34$).

Class 2 had the lowest years of education of the three classes. Class 2 also had a significant estimated lower IQ than Class 1 and this difference was of moderate size ($p$ = .05; Cohen's $d$ = –0.63).

Table 27.

*Descriptive Variables of the Classes from the Long-term Verbal Memory (LTVM) Generalised Growth Mixture Modelling (GGMM) with Baseline Characteristics as a Predictor of Class for the Joint Group*

| | **Participant class** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Class 1** | | **Class 2** | | **Class 3** | | | | | | |
| | **(*n* = 144)** | | **(*n* = 23)** | | **(*n* = 146)** | | | | **Pairwise** | | |
| **Predictor** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***F (df)*** | ***p*** | **comparisons** | ***p*** | **Cohen's *d*** |
| Age (years) | 64.53 | *5.94* | 72.43 | *6.56* | 67.75 | *7.54* | 17.32 *(2, 310)* | <.001*** | 1 *vs.* 2 | <.001*** | –1.26 |
| | | | | | | | | | 1 *vs.* 3 | <.001*** | –0.47 |
| | | | | | | | | | 2 *vs.* 3 | .006** | 0.66 |
| Female (%) | 91.79 | | 34.82 | | 76.75 | | $\chi^2_{(2,\, N = 313)}$ = 43.42 | <.001***[a] | | | Cramer's *V* = 0.372 |
| Education (years) | 14.37 | *3.02* | 12.81 | *3.35* | 13.38 | *2.89* | 5.31 *(2, 310)* | .005** | 1 *vs.* 2 | .053 | –0.49 |
| | | | | | | | | | 1 *vs.* 3 | .014* | –0.335 |
| | | | | | | | | | 2 *vs.* 3 | .665 | –0.183 |
| Estimated Premorbid IQ | 114.03 | *4.610* | 110.74 | *5.78* | 109.86 | *7.62* | 16.41 *(2, 310)* | <.001*** | 1 *vs.* 2 | .052 | –0.63 |
| | | | | | | | | | 1 *vs.* 3 | <.001*** | –0.68 |
| | | | | | | | | | 2 *vs.* 3 | .81 | 0.13 |

[a] Pearson's $\chi^2$ test. * $p < .05$, ** $p. \leq .01$, *** $p < .001$, two-tailed.

*Performance trajectories of classes in the generalised growth mixture modelling with predictors of class membership.* Table 28 shows the parameters (intercepts and slopes) for the three classes in the GGMM incorporating baseline characteristics as predictors. Class 2, the smallest class, had the lowest intercept; Class 3, the largest group, had an intercept value higher than Class 2 but lower than Class 1. Thus the order of classes from lowest to highest initial LTVM performance was Class 2, 3 and 1, respectively. One-way ANOVA revealed that there was a significant difference in intercept values between the classes ($F(2, 308) = 175.45$, $p < .001$). *Post-hoc* comparisons using Tukey HSD indicated that there was a significant difference between all three classes ($p < .001$). Classes 2, 1 and 3 are referred to here as Low, High and Moderate LTVM classes, respectively. In relation to the slope parameters, the Low LTVM class showed no significant change in slope ($p = .60$). In contrast, the slopes of the Moderate and High LTVM classes were significant (both $p < .001$). Figure 11 shows the trajectories of the slopes for each class across 12 months.

Table 28.
*Growth Parameter Estimates for the Classes in the Long-term Verbal Memory (LTVM) Generalised Growth Mixture Modelling (GGMM) with Baseline Characteristics as Predictors of Class for the Joint Group*

| **Variable** | **Estimate** | ***SE*** | ***p.*** |
|---|---|---|---|
| 2: Low LTVM (*n* = 23) | | | |
| Intercept | 5.795 | 1.040 | <.001** |
| Slope | 0.061 | 0.112 | 0.60 |
| 3: Moderate LTVM (*n* = 146) | | | |
| Intercept | 9.669 | 0.266 | <.001** |
| Slope | 0.091 | 0.026 | <.001** |
| 1: High LTVM (*n* = 144) | | | |
| Intercept | 12.397 | 0.192 | <.001** |
| Slope | 0.138 | 0.020 | <.001** |

*Note:* 1, 2, 3 indicates model class assignment in model as per 23 and Table 314.
** $p < .001$, two-tailed.

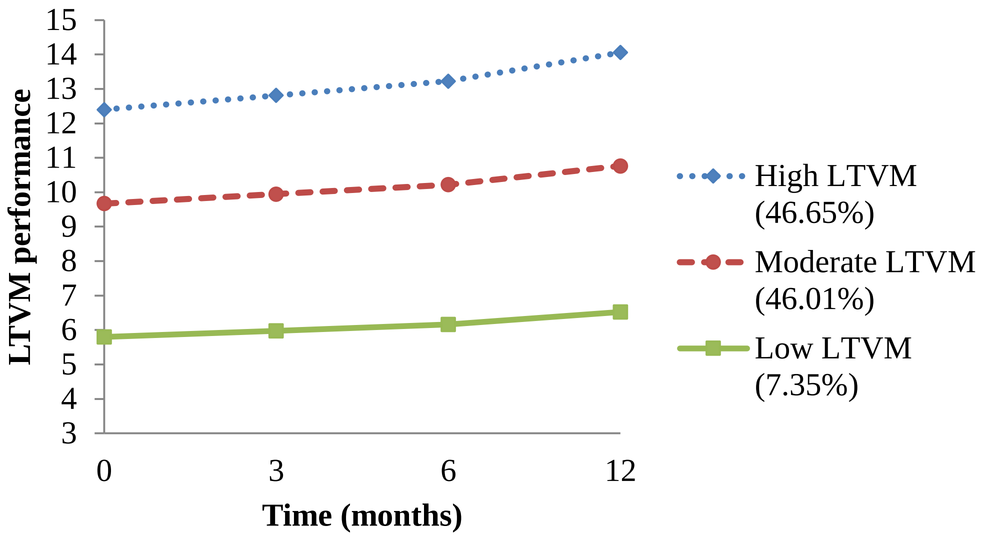


*Figure 11*. Trajectories for the 3-class long-term verbal memory (LTVM) generalised growth mixture modelling (GGMM) across 12 months with baseline characteristics a predictor of class for the joint group ($n$ = 313).

Table 29 and Table 30 show the effects of the predictors – age, sex, education and estimated pre-morbid IQ – on class membership. As can be seen in Table 37, sex significantly discriminated between the Low and Moderate LTVM classes, such that individuals were more likely to be male if they belonged to the Low LTVM class ($p$ = .006). Age, sex and estimated pre-morbid IQ significantly discriminated between the Low and High LTVM classes ($p$ = .002, $p$ < .001 and $p$ = .001, respectively). If a participant was older, male, and had a lower estimated pre-morbid IQ, there was an increased probability of that individual belonging to the Low LTVM class compared with the High LTVM class.

Table 30 shows that age, sex and estimated pre-morbid IQ significantly discriminated between the High and Moderate LTVM classes ($p$ = .001, $p$ = .002 and $p$ < .001, respectively). There was a significant increase in the probability of belonging to the Moderate LTVM class if participants were older (although not as old as the Low LTVM class), male, and had a lower estimated pre-morbid IQ compared with individuals in the High LTVM class. There was also a trend towards an increased probability of individuals belonging to the Moderate LTVM class if they had comparatively less education ($p$ = .07).

Table 29.
*Prediction of Class Membership: Class Comparisons Using the Low Long-term Verbal Memory (LTVM) Class as a Reference Class*

| **Variable** | **Estimate** | ***SE*** | ***p*** |
|---|---|---|---|
| Moderate LTVM ($n$ = 146) | | | |
| Age | –0.092 | 0.052 | .08 |
| Female | –2.069 | 0.751 | .006* |
| Education | 0.144 | 0.111 | .19 |
| Estimated Premorbid IQ | 0.041 | 0.041 | .31 |
| High LTVM ($n$ = 144) | | | |
| Age | –0.174 | 0.056 | .002* |
| Female | –3.566 | 0.861 | <.001** |
| Education | 0.211 | 0.116 | .07 |
| Estimated Premorbid IQ | 0.154 | 0.046 | .001** |

*Note*: Estimate represents the logistic regression coefficient (where negative values represent lower baseline characteristic values for the reference class); sex was coded 1 = female, 2 = male. * $p$ < .01, ** $p \leq$ .001, two-tailed.

Table 30.
*Prediction of Class Membership: Class Comparison Using the High Long-term Verbal Memory (LTVM) Class as a Reference Class*

| **Variable** | **Estimate** | ***SE*** | ***p*** |
|---|---|---|---|
| Moderate LTVM ($n$ = 146) | | | |
| Age | 0.082 | 0.025 | .001** |
| Female | 1.497 | 0.493 | .002* |
| Education | –0.067 | 0.056 | .23 |
| Premorbid IQ | –0.113 | 0.031 | <.001** |

*Note*: Estimate represent logistic regression coefficient (where positive values represent higher baseline characteristic values for the reference class); sex was coded 1 = female, 2 = male. * $p$ < .01, ** $p \leq$ .001, two-tailed. High LTVM *vs.* Low LTVM class comparisons in Table 37.

## Long-term Verbal Memory Performance Summary

In sum, the results presented here show that the conditional LGM for LTVM revealed no significant effect of training status on LTVM performance. The LCGA improved on the LGM, showing three distinct trajectories labelled High, Moderate and Low VM classes, based on their relative baseline performance level (Fandakova et al., 2012; Muthén, online discussion forum correspondence, 2013; Stulz et al., 2010). The GGMM incorporating training status as a covariate showed that experimental participants in the Low LTVM class showed a trend towards improvements in performance trajectories, and the magnitude of the performance difference between controls was large (Cohen's *d* = 1.48). However, the control sample is not considered to be of an adequate size to adequately extrapolate any gains to other populations. The efficacy of training on LTVM is therefore inadequately recovered by the GGMM.

Finally, GGMM showed age, sex and estimated pre-morbid IQ increased the probability that an individual was allocated to the reference Low LTVM class. Specifically, if a participant was older, male, and had a lower estimated pre-morbid IQ, there was an increased probability of that individual belonging to the Low LTVM class compared with the High LTVM class.

RESULTS

# Chapter 9

# Executive Function Performance

## Baseline Performance Differences Between Experimental and Control Participants

Comparisons between the raw baseline EF scores (errors made on the GMLT, CogState Ltd, 2008) of the experimental and control groups were considered. There were no significant differences in scores between the experimental (*M* = 59.40, *SD* = 26.16) and control (*M* = 56.19, *SD* = 18.50) groups; *t*(285) = 0.90, *p* = .37, two-tailed). Both the experimental group and the control groups' EF performances can be considered to fall within the Average range compared with established norms for older individuals (CogState Ltd, 2015).

## Joint Analyses[50]

### Latent Growth Modelling of the Executive Function Scores for the Joint Group

**Unconditional latent growth modelling.** Table 31 shows the growth parameters for the classes (intercepts and slopes) for the unconditional and conditional LGMs for the joint groups' EF performance across the 12-month interval. As noted in the data analysis section, MPlus parameter default options were used. The unconditional models had variances fixed at zero. As can be seen in Table 31, the unconditional LGM (BIC = 8296.803 and ABIC = 8277.775; Table 32), had a significant negative

[50] As noted in the Method chapter, separate analyses were conducted for the control and experimental groups (Appendix 8).

growth in EF errors for the study sample as a whole ($n$ = 285, Estimate = –0.911, $SE$ = 0.17, $p < .001$). This indicated improvement in EF performance across the 12-month period.

Table 31.
*Growth Parameter Estimates for the Conditional and Unconditional Executive Function (EF) Latent Growth Modelling (LGM) for the Joint Group*

| | **Unconditional model** | **Conditional model** |
|---|---|---|
| | **Model estimates (*SE*; *n* = 285)** | **Model estimates (*SE*; *n* = 297)** |
| Intercept | 56.971 (1.140)** | 56.898 (1.35)** |
| Slope | –0.911 (0.17)** | –0.063 (0.28)* |

*Note:* Negative slope estimate value indicates superior experimental group performance compared with controls in conditional model. * $p < .05$, ** $p < .001$.

**Conditional latent growth modelling.** Table 31 also shows that the conditional model (BIC = 8296.054 and ABIC = 8273.854), which allowed variances to vary (*e.g.*, Muthén & Muthén, 2000) and regressed the slope parameter on training status (*i.e.*, experimental *vs.* control groups). There was a significant difference in EF performance trajectories across the 12-month interval between those trained in the ACE program compared with controls ($p = .02$). The EF performance for this model is identified in Figure 12. The largest point of change from baseline for the joint group can be seen at the 12-month assessment point. Overall, whilst the conditional model shows that ACE training effects on EF can be demonstrated in a single performance trajectory of the study population, investigation into the heterogeneity of the study sample EF performance was investigated to better explain performance gains.

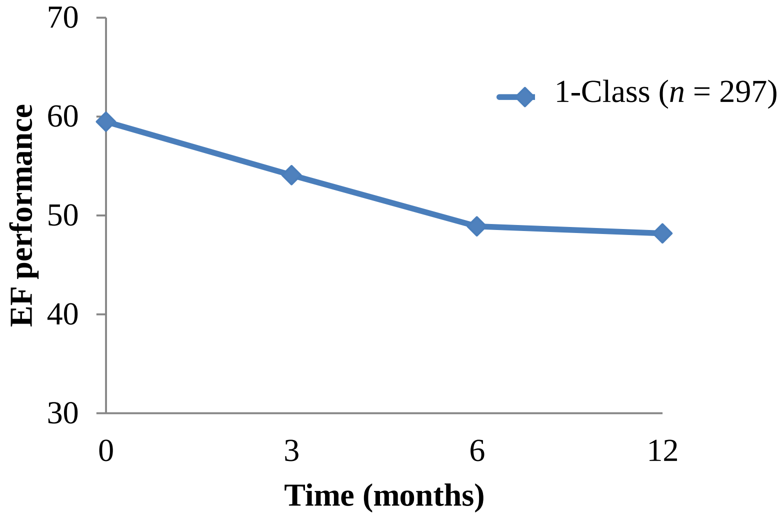


*Figure 12.* Executive Function (EF) latent growth modelling (LGM) estimated growth trajectory across 12 months for the joint group ($n$ = 297).
*Note:* Performance is indicated by number of errors (*i.e.*, fewer errors equates to better performance).

**Latent Class Growth Analysis of the Executive Function Scores for the Joint Groups**

Table 46 shows the model fit indices of the 1- to 3-class unconditional LCGA models conducted to determine the optimal model of EF performance. MPlus parameter default options were used, in addition to variances being fixed at zero. It shows support for a 2-class model. The 2-class solution had the lowest BIC and ABIC values (BIC = 7934.406; ABIC = 7905.864) representing a superior model to the single class model. The 2-class model also demonstrated significant likelihood ratio tests (LMR and adjusted LRT, $p$ = .0004 and $p$ = .0005, respectively), values which were not significant for the 3-class model (LMR and adjusted LRT, $p$ = .16 and $p$ = .17, respectively). The entropy value indicating separation between the two classes was excellent (.929). An independent-samples $t$-test revealed that there was a significant difference at the initial time-point between Class 1 ($M$ = 103.64, $SD$ = 22.55) and Class 2 ($M$ = 51.00, $SD$ = 14.94) in the 2-class model ($t_{285}$ = 19.39, $p$ < .001). The size of the classes can be considered acceptable, whilst the latent class probabilities were 0.987 and 0.952 for Class 1 and Class 2. Conceptually, a 2-class model does not correspond to past studies demonstrating three distinct performance trajectories on global measures of cognitive functioning (*e.g.*, Barnes et al., 2007; as outlined in Chapter 2). However, there have been no other studies to date specifically investigating EF trajectories, thus a 2-class model for EF is entirely plausible. Furthermore, results of this model are also are easily interpretable, including the presence of qualitiatively distinct groups based on initial performance level (Uher et al. 2010). Together with strong statistical support, this 2-class model was therefore accepted.

Table 32.
*Model Fit Indices from the Executive Function (EF) Growth Modelling for the Joint Group*

| Model | BIC | ABIC | LMR *p* | Adjusted LRT *p* | Entropy | Class membership (%) C1 | C2 | C3 |
|---|---|---|---|---|---|---|---|---|
| 1-class | 8296.803 | 8277.775 | | | | 100 | | |
| **2-class** | **7934.406** | **7905.864** | **.0004*** | **.0005*** | **0.929** | **84.51** | **15.49** | |
| 3-class | 7843.417 | 7805.361 | 0.16 | .17 | 0.857 | 18.18 | 74.75 | 7.07 |

*Note:* Bold indicates best fit. BIC = Bayesian information criterion; ABIC = sample-size adjusted Bayesian Information Criterion; LMR = Vuong-Lo-Mendell-Rubin likelihood ratio test; Adjusted LRT = Lo-Mendell-Rubin Adjusted likelihood ratio test. * $p < .001$, two-tailed.

**Descriptive variables of classes for the joint group.** Table 33 presents the mean (*SD*) baseline scores for the two classes for age, sex, number of years of education and estimated pre-morbid (WTAR) IQ scores for the LCGA model incorporating the study sample as a whole. It shows independent-samples *t*-test comparisons of age, years of education and pre-morbid IQ. A $\chi^2$ test for independence, with Yate's Continuity Correction, was used was used to compare the proportions of females in each class. As can be seen in Table 33, there was a significant difference in age and estimated premorbid IQ between the classes ($p$ < .001 and $p$ = .005, respectively). Class 2 had a significantly higher mean age and lower estimated premorbid IQ compared with Class 1; these differences in descriptive variables were of a medium and a small-medium size (Cohen's $d$ = –0.60 and 0.38, respectively). Despite a significant difference in estimated premorbid IQ between the classes, after rounding the estimated IQ scores to the nearest whole number, both classes had a High Average premorbid IQ score.

Table 33.
*Descriptive Variables of the Classes from the Executive Function (EF) 2-Class Growth Model for the Joint Group*

| | **Participant class** | | | | | |
|---|---|---|---|---|---|---|
| | **Class 1 (*n* = 251)** | | **Class 2 (*n* = 46)** | | | |
| | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***t (df)*** | ***p*** |
| Age | 65.87 | 6.75 | 70.25 | 7.74 | 15.65 (295) | <.001** |
| Female (%) | 80.14 | | 78.33 | | $\chi^2$ (1, $N$ = 235) = 9.34 | 0.84 [a] |
| Education (years) | 14.01 | 3.00 | 13.19 | 3.05 | 2.89 (295) | .09 |
| Estimated Premorbid IQ | 112.48 | 5.83 | 109.56 | 9.04 | 7.95 (295) | .005* |

[a] $\chi^2$ test of independence with Yate's Continuity Correction. * $p$ < .01, ** $p$ < .001, two-tailed.

*Performance trajectories of classes for the joint group.* Table 34 presents the parameter estimates for the two classes identified. Class 1, the larger class and had the lowest intercept of the two classes, indicating superior EF performance (*i.e.*, fewer errors) than Class 2. As previously noted, there was a significant difference between the intercept values ($p$ < .001). Classes 1 and 2 are therefore referred to as High and Low EF classes, respectively. When considering the slope parameters, there was a significant positive slope for both the High and Low EF classes ($p$ < .001 and $p$

= .004, respectively). Figure 13 demonstrates the trajectories across 12 months for each class. These results suggest that EF performance for the Low and High EF classes improved across the 12-month follow-up period.

Table 34.
*Growth Parameter Estimates for the Classes in the Executive Function (EF) 2-Class Model for the Joint Group*

| **Variable** | **Estimate** | ***SE*** | ***p*** |
|---|---|---|---|
| 2: Low EF (*n* = 46) | | | |
| Intercept | 99.492 | 4.893 | <.001** |
| Slope | –1.841 | 0.642 | .004* |
| 1: High EF (*n* = 251) | | | |
| Intercept | 49.745 | 1.168 | <.001** |
| Slope | –0.771 | 0.093 | <.001** |

*Note:* 1, 2 indicates model class assignment in model as per Table 33. * $p$ < .01, ** $p$ < .001, two tailed.

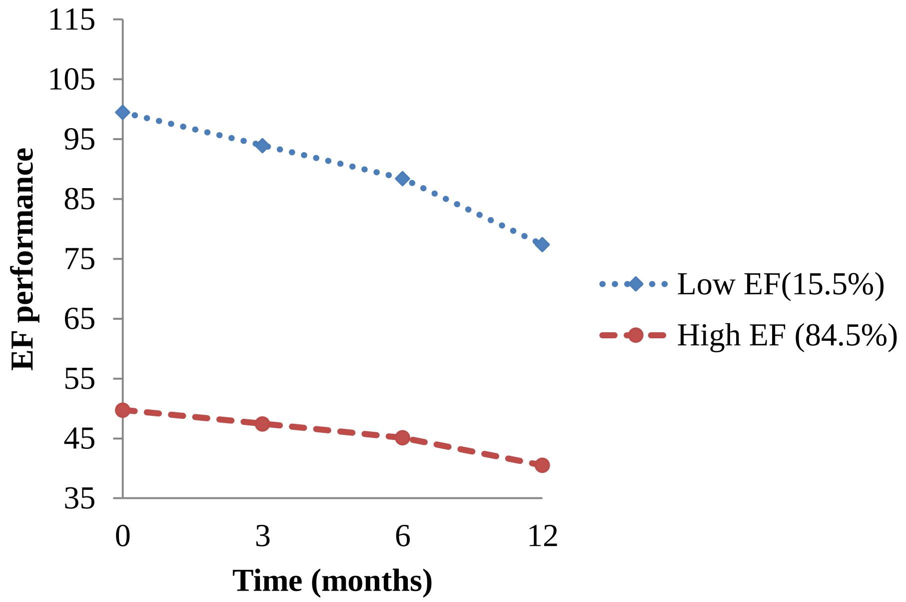


*Figure 13.* Trajectories of the executive function (EF) latent class growth analysis (LCGA) classes across 12 months for the joint group ($n$ = 297).
*Note.* Performance is indicated by number of errors (*i.e.*, fewer errors equates to better performance).

**Generalised Growth Mixture Modelling with Training Status as a Predictor of Executive Function for the Joint Group**

Table 35 shows the various model fit indices for GGMM to determine the optimal final model of training effects on EF performance for the joint groups. Again, MPlus parameter default options were used. As was previously highlighted in the unconditional analyses, model assessment was conducted incrementally from 1 to 3 classes. The model of best fit was selected from these statistical fit indices as well as conceptual considerations. Table 35 shows that a 2-class model was supported, as it was for the unconditional model. The 2-class GGMM model (BIC = 7860.17 and ABIC =7812.6) also had significant LMR and adjusted LRT values ($p$ = .009 and $p$ = .01, respectively). An independent-samples $t$-test revealed that there was a significant difference at the initial time-point between Class 1 ($M$ = 102.59, $SD$ = 22.59) and Class 2 ($M$ = 50.76, $SD$ = 14.74) in the 2-class model ($t(285) = 19.56$, $p < .001$). The entropy value indicated that separation between the two classes was excellent (0.932). Class 1 contained only 16.16% of the total cohort, which was considered acceptable. Furthermore, the model demonstrated latent class probabilities of 0.954 and 0.989 for Classes 1 and 2, respectively. As previously outlined in reference to the unconditional LCGA for EF performance, conceptually, a 2-class model did not correspond to past studies demonstrating three distinct performance trajectories on global measures of cognitive functioning (*e.g.*, Langbaum et al., 2009; Willis et al. 2006). A 2-class model was accepted, however given the novelty of the present study in investigating EF trajectories it was accepted and considered reasonable. Results of this model were also easily interpretable, with the presence of qualitiatively distinct groups based on initial performance level (*e.g.*, Uher et al., 2010). Together with strong statistical support, this model was therefore accepted.

Table 35.
*Model Fit Indices from the Executive Function (EF) Generalised Growth Mixture Modelling (GGMM) Incorporating Training Status as a Predictor for the Joint Group*

| **Model** | **BIC** | **ABIC** | **LMR *p*** | **Adjusted LRT *p*** | **Entropy** | **Class membership (%) C1** | **C2** | **C3** |
|---|---|---|---|---|---|---|---|---|
| 1-class | 8296.05 | 8273.85 | | | | 100 | | |
| **2-class** | **7860.17** | **7812.6** | **.009**** | **.01*** | **0.932** | **16.162** | **83.838** | |
| 3-class | 7935.97 | 7901.08 | .20 | .21 | 0.857 | 7.071 | 17.845 | 75.084 |

*Note:* Bold indicates best fit. BIC = Bayesian information criterion; ABIC = sample-size adjusted Bayesian Information Criterion; LMR = Vuong-Lo-Mendell-Rubin likelihood ratio test; Adjusted LRT = Lo-Mendell-Rubin Adjusted likelihood ratio test. * $p$ = .01, ** $p$ < .01, two-tailed.

**Descriptive variables of classes in the generalised growth mixture modelling with training status as a predictor.** Table 36 presents the mean (*SD*) scores for the two classes for age, sex, number of years of education and estimated pre-morbid (WTAR) IQ scores for the GGMM incorporating training status (*i.e.*, experimental *vs.* control group) as a predictor of EF performance across the 12-month interval. The table shows independent-samples *t*-tests for comparison of age, years of education and pre-morbid IQ. A $\chi^2$ test for independence, with Yate's Continuity Correction, was used to compare the proportions of females in each class. As can be seen in the table, the baseline descriptive variables of the classes in the GGMM were very similar to those in the LCGA model (Table 33). Class 1 had a significantly higher mean age ($M$ = 69.64 years) than Class 2 ($M$ = 65.95 years; $p$ = .001) and was of medium size (Cohen's $d$ = 0.49). In addition, Class 1 had a significantly lower estimated IQ, a difference which was also of medium size ($p$ = .002, Cohen's $d$ = –0.42). Class 1 would be considered to have an Average estimated premorbid IQ, whilst Class 2 had a High Average IQ score.

Table 36.
*Descriptive Variables of the Classes from the Executive Function (EF) Generalised Growth Mixture Modelling (GGMM) with Training Status as a Predictor for the Joint Group*

| | Participant class | | | | | |
|---|---|---|---|---|---|---|
| | **Class 1 (*n* = 48)** | | **Class 2 (*n* = 249)** | | | |
| | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***t (df)*** | ***p*** |
| Age | 69.64 | *8.16* | 65.95 | *6.71* | 3.36 *(295)* | .001** |
| Female (%) | 79.20 | | 79.90 | | $\chi^2_{(1, N = 297)}$ <.001 | .99[a] |
| Education (years) | 13.22 | *3.08* | 14.01 | *3.00* | –1.66 *(295)* | .10 |
| Estimated Premorbid IQ | 109.40 | *8.88* | 112.53 | *5.82* | –3.08 *(294)* | .002* |

[a] $\chi^2$ test of independence with Yate's Continuity Correction. * $p$ < .01, ** $p$ = .001, two-tailed.

**Performance trajectories of classes in the generalised growth mixture modelling with training status as a predictor.** Table 37 shows the parameters (intercepts and slopes) for the two classes in the GGMM model assessing effects of the ACE training, comparing experimental and control groups. Class 1 was the smaller class and had the higher intercept (indicating more EF errors) compared with Class 2. Thus Class 1 had a poorer EF performance than Class 2. As previously noted, there was a significant difference between the initial values of the two groups ($p$ < .001). Classes 1 and 2 are referred to here as the Low and High EF classes, respectively. In relation to the slope parameter, the results indicate that membership of the treatment group significantly predicted increased performance trajectories for the experimental participants in the Low EF ($p$ = .03), compared with controls. The experimental participants in the High EF did not, however, show a significant profit from the intervention compared with controls ($p$ = .40).

Table 37.
*Growth Parameter Estimates for the Classes in the Executive Function (EF) Generalised Growth Mixture Modelling (GGMM) with Training Status as a Predictor for the Joint Group*

| **Variable** | **Estimate** | ***SE*** | ***p*** |
|---|---|---|---|
| 1: Low EF (*n* = 48) | | | |
| Intercept | 98.632 | 4.02 | <.001** |
| Slope | –1.832 | 0.851 | .03* |
| 2: High EF (*n* = 249) | | | |
| Intercept | 49.537 | 1.004 | <.001** |
| Slope | –0.172 | 0.204 | .40 |

*Note:* Negative slope estimate values indicate superior performance of experimental group compared with controls; 1, 2 indicates model class assignment in model as per Table 36 and Table 37. * $p < .05$, ** $p < .001$, two-tailed.

Figure 14 shows the trajectories for each of the classes in the model (lower values indicate fewer errors and thus better performance). The left panel demonstrates the estimated trajectories, whilst the right panel separates the control and experimental constituents of each of the two classes and their observed trajectories across the 12-month time interval. Blue shaded lines represent the Low EF class and red shaded lines represent the High EF class. As can be seen in the left panel, the number of EF errors decreased (*i.e.*, performance increased) across the 12-month follow-up for both EF classes and errors were at their lowest for both classes at the 12-month interval. The Low EF class performed at a Low Average level at baseline compared to norms, whilst the High EF class performed at an Average level compared to norms (CogState Ltd, 2015). Importantly, as noted in Table 37, there was a significant difference in the slopes of the experimental and control participants in the Low EF class. The EF performance difference between the groups' *slopes* in the Low EF class was of a large magnitude (Cohen's $d$ = 2.23; Valenzuela & Sachdev, 2006a). There was no significant difference between the groups' slopes in the High EF class ($p$ = .40)

The right panel shows some improvement in the EF performance of both the experimental and control participants' in the Low EF class. When the two groups' estimated slopes were compared, however, as previously noted, the experimental and control groups' performances were significantly different, with the experimental group showing superior performance compared with controls (see Table 37). There was also a difference between the experimental and control groups at the 12-month mark that was of a small size (Cohen's $d$ = 0.26). Whist there was a difference in EF performance at baseline between the experimental and control participants in the Low EF, this was not statistically significant ($t(39) = 2.23$, $p$ = .06).

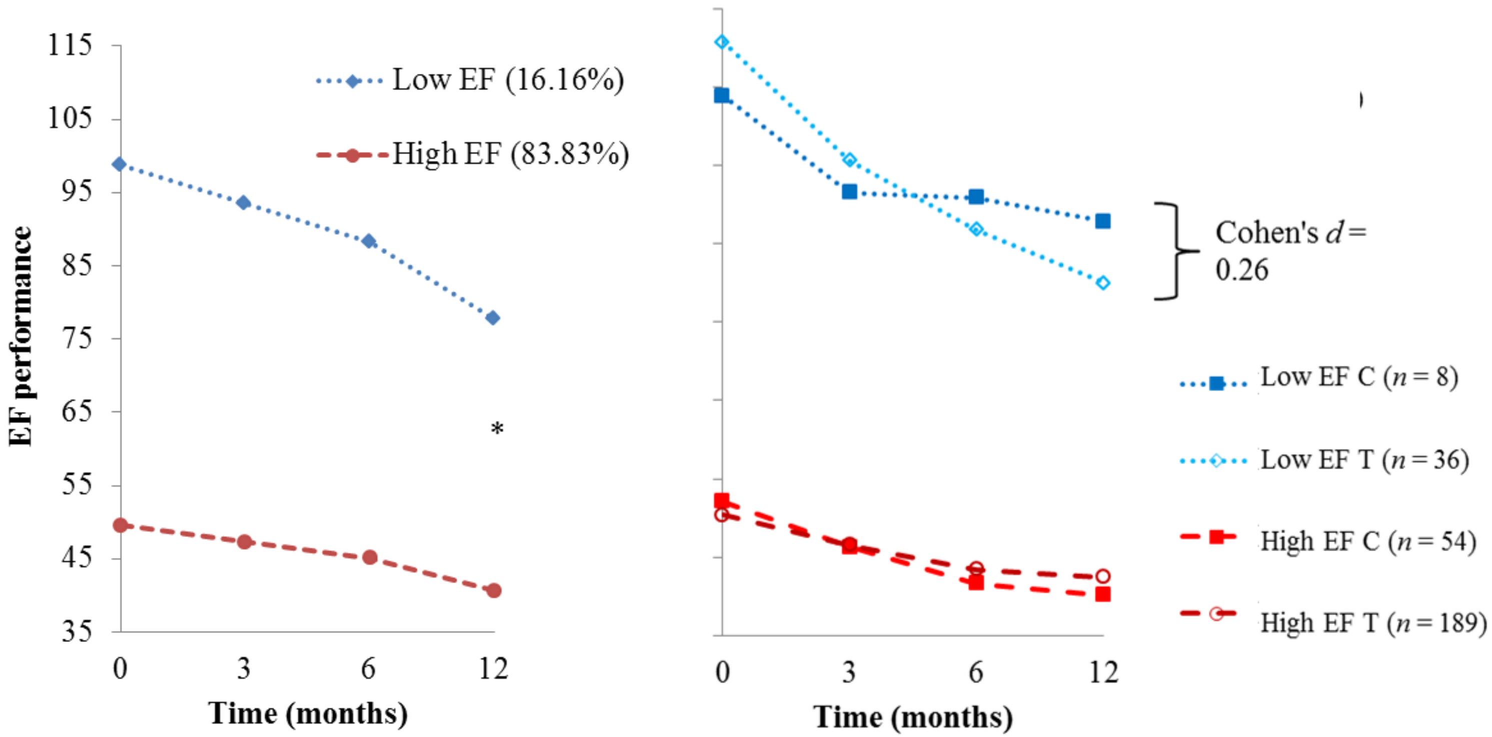


*Figure 14*. Estimated (left panel) and observed (right panel) trajectories across 12 months for the 2 classes of the executive function (EF) generalised growth mixture model (GGMM) incorporating training status as a predictor for the joint group ($n$ = 297).
*Note*: E = experimental group; C = control group. * $p$ = .01, two-tailed, Cohen's $d$ = 2.23 (for slope comparisons; left panel). Performance is indicated by number of errors (*i.e.*, fewer errors equates to better performance).

Both Figure 14 and Table 38 show the effect sizes (Cohen's *d*) of comparisons of the observed EF performance between the initial observed performance and the 12-month follow-up for the experimental and control participants. This highlights the magnitude of changes across the 12-month interval from baseline (Valenzuela & Sachdev, 2006a). As noted, the effect size for the Low EF was small (Cohen's $d$ = 0.26). There was also a small effect of treatment for the High EF class (Cohen's $d$ = 0.34), however again it should be noted that these effects were not statistically significant (Table 37).

Table 38.
*Effect Sizes for the Baseline to 12-month Comparisons for Experimental and Control Groups in the 3-Class Executive Function (EF) Generalised Growth Mixture Modelling (GGMM) Incorporating Training Status as a Predictor*

| | **Group** | | |
|---|---|---|---|
| | **Experimental** | **Control** | **Effect size difference** |
| Low EF | 0.545 | 0.805 | 0.26 |
| High EF | 0.538 | 0.882 | 0.344 |

*Note:* Effect size, Cohen's *d*, using Morris and DeShon's (2002) equation 8 to correct for dependence between means.

Importantly, the overall results of this generalised model for EF suggest that participating in the ACE program training predicted both statistically significant improvement of a meaningful magnitude for the EF performance trajectories for the experimental group in the Low EF class. Specifically, ACE-trained individuals demonstrated improvements in EF performance both when their estimated performance trajectories are compared with those of the controls and across the 12-month follow-up period. However, of note, there were only eight controls participants in the Low EF class to which the 36 treatment group participants were compared. Results were therefore interpreted cautiously.

## Generalised Growth Mixture Modelling Predicting Class Membership with Baseline Characteristics

As previously outlined, predictors of class membership – age, sex (1 = female, 2 = male), years of formal education and estimated pre-morbid IQ – were explored in the second GGMM. Most MPlus default parameters were used; however, starts were increased from the initial stage random sets of 10 values to 500 and residual variances were held at zero to address non-convergence, as recommended (Muthén & Muthén, 1998–2010; Petras & Masyn, 2010). Table 39 shows the various model fit indices for these models, including 1- to 3-classes. As can be seen, whilst the 3-class model had the lowest BIC and ABIC values, the LMR and adjusted LRT values were not significant ($p$ = .27). The 2-class model had LMR and adjusted LRT values approaching significance ($p$ = .06). This model also had BIC and ABIC values that improved on the unconditional 2-class model (GLCGA predicting class membership: BIC = 7915.53, ABIC = 7874.302 *vs.* LCGA BIC = 7934.406, ABIC = 7905.864).

The entropy value indicated that separation between the two classes was excellent (0.93). Class 1 contained 15.02% of the total cohort, which, as previously noted, was acceptable. Furthermore, the model demonstrated latent class probabilities of 0.948 and 0.986 for Classes 1 and 2, respectively. As previously outlined for the earlier GGMM investigating treatment status (Table 36), the effects on the performance trajectories showed that, theoretically, a 2-class model for EF was plausible. As such the 2-class model was accepted. There was also conceptual support for the 2- class model, based on the previous GGMM of EF in the present study.

Table 39.
*Model Fit Indices from the Executive Function (EF) Generalised Growth Mixture Modelling (GGMM) Incorporating Training Status as a Predictor of Class for the Joint Group*

| **Model** | **BIC** | **ABIC** | **LMR *p*** | **Adjusted LRT *p*** | **Entropy** | **Class membership (%) C1** | **C2** | **C3** |
|---|---|---|---|---|---|---|---|---|
| 1-class [a] | 14438.469 | 14394.065 | | | | 100 | | |
| **2-class** | **7839.211** | **7775.785** | **0.06** | **0.06** | **0.927** | **15.203** | **84.797** | |
| 3-class | 7935.97 | 7901.08 | .27 | .27 | 0.841 | 70.27 | 22.635 | 7.095 |

*Note:* [a] Predictor effects were modelled on the slope parameter for the 1-class model; in all other models, predictor effects were exclusively modelled on class membership. Bold indicates best fit. BIC = Bayesian information criterion; ABIC = sample-size adjusted Bayesian Information Criterion; LMR = Vuong-Lo-Mendell-Rubin likelihood ratio test; Adjusted LRT = Lo-Mendell-Rubin Adjusted likelihood ratio test.

**Descriptive variables of classes in the generalised growth mixture modelling with predictors of class membership.** Table 40 shows the mean (*SD*) baseline age, sex, number of years of education and pre-morbid (WTAR) IQ scores for the three GGMM classes. It also shows the results of the appropriate statistics (independent-samples *t*-tests for comparison of age, years of education and pre-morbid IQ and a $\chi^2$ test for independence, with Yate's Continuity Correction, to compare the proportions of females in each class). The results indicated significant class differences for age and estimated premorbid IQ ($p$ < .001 and $p$ = .005, respectively). Class 1 was significantly older and had a significantly lower estimated premorbid IQ than Class 2. These differences were of a large and small-medium size, respectively (Cohen's $d$ = 0.62 and 0.38, respectively).

Table 40.
*Descriptive Variables of the Classes from the Executive Function (EF) Generalised Growth Mixture Modelling (GGMM) with Baseline Characteristics as Predictors of Class for the Joint Group*

| | **Participant class** | | | | | |
|---|---|---|---|---|---|---|
| | **Class 1 (*n* = 45)** | | **Class 2 (*n* = 251)** | | | |
| | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***t (df)*** | ***p*** |
| Age | 70.36 | *7.79* | 65.87 | *6.75* | 4.01 *(294)* | <.001** |
| Female (%) | 80.00 | | 80.10 | | $\chi^2$ (1, *N* = 296) | .99 |
| Education (years) | 13.19 | *3.084* | 14.01 | *3.00* | –1.69 *(294)* | .09 |
| Estimated Premorbid IQ | 109.56 | *9.04* | 112.48 | *5.83* | –2.82 *(294)* | .005* |

* $p < .01$, ** $p < .001$, two-tailed.

*Performance trajectories of classes in the generalised growth mixture modelling with predictors of class membership.* Table 41 shows the parameters (intercepts and slopes) for the three classes in the GGMM incorporating baseline characteristics as predictors. Class 1 was the smaller of the two classes and had the highest intercept (*i.e.*, the greatest errors thus poorest performance). An independent-samples *t*-test showed a significant difference in intercept values between the classes model ($t(284) = 19.01$, $p < .001$). Classes 1 and 2 are referred to here as Low and High EF, respectively. Figure 15 shows the trajectories of the slopes for each class across 12 months. In relation to the slope parameters, both the Low and High EF classes showed significant change in slope ($p = .01$ and $p < .001$, respectively).

Table 41.
*Growth Parameter Estimates for the Classes in the Executive Function (EF) Generalised Growth Mixture Modelling (GGMM) with Baseline Characteristics as Predictors of Class for the Joint Group*

| **Variable** | **Estimate** | ***SE*** | ***p*** |
|---|---|---|---|
| 1: Low EF (*n* = 45) | | | |
| Intercept | 98.86 | 5.27 | <.001** |
| Slope | –1.763 | 0.683 | .01* |
| 2: High EF (*n* = 251) | | | |
| Intercept | 49.805 | 1.295 | <.001** |
| Slope | –0.771 | 0.092 | <.001** |

*Note:* 1, 2 indicates model class assignment as per Table 40. * *p* = .01, ** *p* < .001, two-tailed.

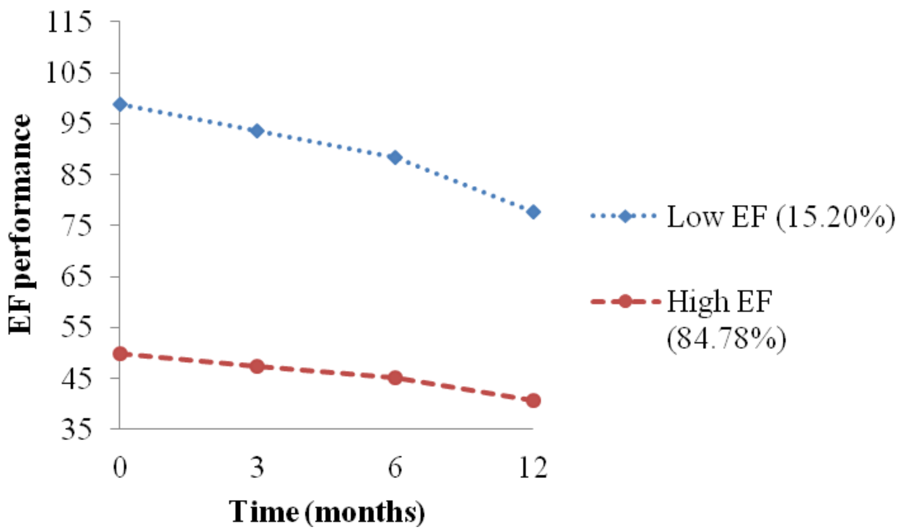


*Figure 15.* Trajectories for the 2-class executive function (EF) generalised growth mixture modelling (GGMM) across 12 months with baseline characteristics a predictor of class for the joint group (*n* = 297).

Table 42 shows the effects of the predictors age, sex, education and estimated pre-morbid IQ on class membership. As can be seen, age and estimated pre-morbid IQ significantly discriminated between the Low and High EF classes (*p* = .005 and *p* = .008, respectively). If a participant was older, and had a lower estimated pre-morbid IQ, there was an increased probability of that individual belonging to the Low EF class compared with the High EF class.

Table 42.
*Prediction of Class Membership: Class Comparison Using the Low Executive Function (EF) Class as a Reference Class*

| **Variable** | **Estimate** | ***SE*** | ***p*** |
|---|---|---|---|
| High EF (*n* = 251) | | | |
| Age | 0.109 | 0.039 | .005* |
| Female | –0.006 | 0.498 | .99 |
| Education | –0.035 | 0.067 | .61 |
| Estimated Premorbid IQ | –0.089 | 0.033 | .008* |

*Note*: Estimate represents the logistic regression coefficient (where negative values represent lower baseline characteristic values for the reference class); sex was coded 1 = female, 2 = male. * $p < .01$, two-tailed. Low EF *vs.* Moderate EF class comparisons in Table 41.

## Executive Function Results Summary

In sum, the results presented here show that the conditional LGM analysis on the EF performance revealed a significant difference in EF performance trajectories across the 12-month interval between experimental participants compared with controls. However, LCGA showed performance was better represented by heterogeneous cognitive performances. Specifically, a two class model with High EF and a Low EF performance classes. The classes were labelled based on their baseline performance level, relative to their performance standing in relation to the study cohort who executed the task (Fandakova et al., 2012; Muthén, online discussion forum correspondence, 2013; Stulz et al., 2010). The GGMM incorporating training status as a covariate showed a statistically significant and large effect of training on EF trajectory performance slope across 12-months for the Low EF group. However, whilst training effects are plausible, due to the small number of controls ($n$ = 8) to which the experimental participants were compered in the Low class, generalisability of results were interpreted cautiously. Whilst the High EF group improved, there was no statistically significant difference between the experimental participants and the controls. Finally, GGMM showed age and estimated pre-morbid IQ increased the probability that an individual was allocated to the reference Low EF class. Specifically, if a participant was older and had a lower estimated pre-morbid IQ, there was an increased probability of that individual belonging to the Low EF class compared with the High EF. Sex and education did not predict Low EF class membership class.

# Chapter 10

# Discussion

This study had three aims. The first aim was to investigate heterogeneity in cognitive training effects versus controls in longitudinal cognitive trajectories using GGMM. Specifically, the study aimed to explore performance trajectories across a 12 month follow-up period following participation in the ACE program. The second aim was to examine specific and generalised effects of training. Specifically, to examine VM performance following training representing more specific effects of training, as well as LTVM and EF performance trajectories, representing more generalised effects.[51]

The third aim was to identify individual baseline characteristics, such as age, sex and proxies for CR (education and estimated premorbid IQ) that predict and therefore distinguish the different the three cognitive performance trajectories (VM, LTVM and EF) for the study sample.

## Hypotheses

Three hypotheses were made. The first was that there would be distinguishable inter-individual differences in cognitive performance following training compared to controls. The second hypothesis was that hypothesis one would apply across all cognitive domains measured: VM, LTVM and EF performance trajectories, representing specific (VM) and more generalised effects (LTVM and EF) of training on cognitive performance. The third hypothesis was that baseline characteristics age,

[51] VM and LTVM performance was explored and measured using Trial 5 and the delayed recall measure of the RAVLT, respectively (Rey, 1941; 1964). Executive function, measured on the GMLT (CogState Ltd, 2015).

sex, and proxies for CR, education and estimated premorbid IQ would be found as predictors of class membership. Specifically younger females who had a higher estimated premorbid IQ and education would show greatest improvement in the 12-month follow-up trajectories.

## Statistical Analyses Implemented

The GGMM techniques employed to address the above mentioned aims and hypotheses (Muthén, 2002; Muthén & Shedden, 1999). Generalised growth mixture models consider different classes as representing meaningful strata and enable consideration of data as representing reality (see Kreuter & Muthén, 2007; Nagin & Odgers, 2010), or sufficiently so to guide theory development and/or clinical decision-making (Gueorguieva et al., 2007; Kreuter & Muthén, 2007).

Analyses also included LGM and another group-based growth modelling technique LCGA. In addition, as is standard protocol when conducting growth models, the control and treatment groups were regarded as different populations with different growth trajectories. These models were included as a demonstration of model-building procedures and for refining model specification (*i.e.*, no conclusions were drawn from these models). Thus the models were presented separately in Appendices (Huang et al., 2010; Jung & Wickrama, 2008; Nagin, 1999).

It is believed that this is the first study in the older adult literature to assess inter-individual differences in cognitive training-related changes in memory and EF using GGMM.

### Support for Models Utilised

From a GGMM perspective the models produced were considered acceptable and valid growth mixture models demonstrating heterogeneity of performance following training. Acceptability of models was indicated by a number of established GGMM guidelines (Jung & Wickrama, 2008). These include sample size (*e.g.*, Attix et al., 2008; Muthén, 2004; Stulz et al., 2010), a number of model fit indices (*e.g.*, BIC; ABIC; LMR; Adjusted LRT), and class size, with all classes containing above 1% of the sample and being qualitatively distinct (based on entropy values and significant *t* tests; Jung & Wickrama, 2008). In addition, there was consistency of class enumeration across models in the present study. There were three classes for the VM and LTVM domains and two classes for the EF domain, for both the LCGA and the GGMM. Similarly there was consistency of class enumeration across models in other studies investigating memory performance thereby demonstrating conceptual support (*e.g.*, Langbaum et al., 2009; Gross et al., 2012). Langbaum and colleagues (2009) examined differential performance trajectories of participants based on training responsiveness on a variety of memory tasks in the memory training arm of the ACTIVE study ($n$ = 703). As mentioned in Chapter 4, they identified three distinct groups: the 'HVLT class', which can be considered a high performing class; the RAVLT class consisting of individuals with a slightly lower performance improvement; and a third group defined as a "low-level response class" (p.15). Thus, like the current study, Langbaum and colleagues importantly showed three distinct classes following training, when considering inter-individual differences in memory performance when utilising more sophisticated statistical techniques. The congruence between the number of classes within the models in the present study, and comparisons with other studies supports the indication that these data represent meaningful 'real world' strata for memory performance. The two classes identified for the EF measure in this study are a novel contribution to the literature with regard to class enumeration, given that there have been no other known studies to examine EF performance trajectories following training.

Together this identification of inter-individual differences, particularly for EF performance, provides useful a priori information to guide future explorations of cognitive performance trajectories following training. With knowledge of the number of classes, researchers can move beyond the potential ambiguity of empirical considerations of optimal growth model selection by adding more certainty to model selection when similar domains are explored.

## Hypotheses 1 and 2: Heterogeneity of Longitudinal Cognitive Performance Trajectories Following Training – Specific and Generalised

### Single Group Cognitive Trajectory Performance

To investigate the first hypothesis of the study, an initial exploration by LGM for the VM controls revealed that the normative group trajectory showed no significant growth across the 12-month period. The LTVM controls demonstrated a significant growth, as did the EF control group. The LGM of the VM, LTVM and EF treatment groups all showed significant positive growth across time. The treatment group findings could suggest that the ACE program was effective when considering one common mean growth curve. That said, no conclusions were drawn from these analyses, given that direct comparisons between the classes were not made. Whilst the control trajectories can be considered to be normative trajectories in the absence of intervention (Muthén et al., 2002), these two separate LGM models represent only an approximation of differences between the treatment and control groups (Muthén, personal correspondence, January 13, 2009). As noted, they were conducted to build and refine models (Huang et al., 2010; Jung & Wickrama, 2008; Nagin, 1999). The conditional joint LGM analyses – *i.e.*, the model including both the control and treatment group individuals, which included training status as a covariate (*e.g.*, Muthén, 2002) – was considered the final model from which to draw conclusions regarding training.

The conditional LGMs for the VM and LTVM outcome measures revealed no ACE treatment effect on VM and LTVM performance. These findings, therefore, supported the current study's hypothesis that training effects could not be demonstrated in a single performance trajectory of the study population. It revealed that the conventional LGM approach, that assumes performance can be modelled in a single growth trajectory, does not adequately approximate the groups' memory performance. In contrast, the conditional LGM analysis on the EF performance of the cohort revealed a significant difference in EF performance trajectories across the 12-month interval between those trained in the ACE program compared with controls. This finding was counter to what was expected, given that past research in cognitive training suggests that a single trajectory may not adequately demonstrate training effects, given inter-individual responsiveness to training (Ball et al., 2002; Martin et al., 2011; Park et al,, 2007; Terrera et al., 2010). Nonetheless further more fine-grained analysis with GBM revealed a better model fit when heterogeneity was considered, as will be discussed below.

**Heterogeneity of Cognitive Performance Trajectories Following Training**

Next, LCGA was conducted to further address the first hypothesis of the study, that cognitive performance is best modeled with distinct groups (*i.e.*, heterogeneity of performance). Again, as a preliminary statistical step, models for the control and treatment groups were created. The LCGA for the two groups revealed statistical improvement of the models produced in comparison to the LGM for all of the outcome measures (see Tables 6, 19 and 32). Three class models – including High, Moderate and Low performance trajectories – were identified for the VM and LTVM outcome measures. The classes were labelled based on their baseline performance level on the respective RAVLT outcome measures, relative to the cohort (Fandakova et al., 2012; Muthén, online discussion forum correspondence, 2013; Stulz et al., 2010). In the LCGA, the control trajectories were considered to represent normative heterogeneous trajectories in the absence of intervention (Muthén et al., 2002). In addition, the number of control and treatment group classes was the same, which suggested that the intervention did not influence class membership (Muthén et al., 2002). However, as for the unconditional LGM, no conclusions were drawn from these models with regard to treatment effects, as they only represented an approximation of the differences between the control and treatment groups' multiple class trajectories; the results were used in the iterative model-building process (Huang et al., 2010; Jung & Wickrama, 2008; Nagin, 1999; Muthén personal correspondence, January 13, 2009). The joint LCGAs improved on the single class model, also revealing three distinct trajectories labelled High, Moderate and Low VM and LTVM classes. Once again trajectories were labelled based on their relative baseline performance level (Fandakova et al., 2012; Muthén, online discussion forum correspondence, 2013; Stulz et al., 2010). As mentioned, the conditional EF LGM demonstrated a single EF performance trajectory in which there was a significant difference between those trained in the ACE program compared with controls. Nonetheless, a LCGA was carried out to determine if performance was better represented by heterogeneous cognitive performances. An improved model based on two key empirical criteria – the BIC and the ABIC values – was revealed. Based on these values, the optimal model was deemed to have two distinct classes, a High EF and a Low EF performance class. The classes were also labelled based on their baseline performance level, relative to their performance standing in relation to the study cohort who executed the task (Fandakova et al., 2012; Muthén, online discussion forum correspondence, 2013; Stulz et al., 2010).[52].

The final model conducted for each of the three outcome measures was the GGMM, incorporating cognitive training as a covariate. It addressed the first hypothesis of the study, to investigate inter-individual differences in training responsiveness. The GGMM demonstrated that individuals showed heterogeneity in the effects of training, as predicted. The GGMM revealed demonstrable gains in some trained individuals.

**Verbal memory and long-term verbal memory.** Specifically, in relation to intervention effects on VM and LTVM memory, the generalised growth models showed that the treatment group for the Low performance classes had statistically significant long-term gains in cognitive performance trajectories for VM performance. The magnitude of the experimental effect on the slopes was large (Valenzuela & Sachdev, 2006a), with Cohen's $d$ = 4.48 and 2.23, respectively. The Low LTVM trended towards significance and the magnitude of the experimental effect on the performance trajectory was large (Cohen's $d$ = 1.48). However, for both the VM and LTVM, comparisons between the experimental groups in the Low class were made with only two controls. As previously noted it is statistical convention in growth modelling to focus on the class number in total, and more generally whether they were

[52] It should be noted that whilst this outcome measure indicated an individual's raw error score on CogState Ltd's GMLT, they were labelled according to consideration of successful performance of the task. That is, the Low EF class had a higher error score, indicating lower EF performance and *vice versa*.

statistically and conceptually sound (*e.g.* they contained at least 1% of the sample and balanced goodness of fit with parsimony; Jung & Wickrama, 2008). Nonetheless, the present study considers the models as inadequately predicting specific (VM) and generalised (LTVM) training gains due to the low number of controls to which the treatment groups were compared. Thus the training results cannot be extrapolated to other populations and hypothesis two was not supported with these results. This has implications regarding the use of GGMM itself and highlights that caution must be taken when utilising GGMM models to draw conclusions, despite apparent statistical acceptability.

The present memory findings are consistent with some investigations showing a lack of robust and replicable effects of cognitive training from trained to untrained skills in the *memory* domain (Martin et al., 2011; Owen et al., 2010; Papp et al., 2009; Valenzuela & Sachdev, 2009; Verhaeghen et al., 1992). Owen and colleagues (2010) conducted a six-week online training study ($n$ = 11,430) in which participants were trained several times each week on cognitive tasks designed to improve memory (as well as reasoning, planning, visuospatial skills and attention). No clear evidence was found that, compared to controls, the experimental group demonstrated training effects. This was the case even when the memory task (a classic parlour game in which players have to remember the locations of objects on cards), was closely related to the benchmarking task (a paired-associates learning task; PAL).

Previous meta-analyses have also reported limited demonstration of efficacy of training (*e.g.*, Martin et al., 2011; Papp et al., 2009; Valenzuela & Sachdev, 2009). For example, Valenzuela and Sachdev (2009) conducted a meta-analysis of seven randomised control trials (RCTs; $N$ = 3,194) that tested the longitudinal (>3 months) neuropsychological performance effects of cognitive training programs, including memory. Generalised effects were only reported in two of the seven studies.

As in the present study, past research has also highlighted issues with the control groups (Ball et al., 2002; Langbaum et al., 2009; Martin et al., 2011; Owen et al., 2010). The number of controls in Owen and colleagues' (2010) study was not well matched with the experimental participants. In their case, this was due to the retention rate being much lower in the control group. Moreover, the active controls' memory improved, with no formal memory training at all. Martin and colleagues (2011) also reported improvement in control conditions across studies in their meta-analysis. This will be further discussed below and in the limitations section.

**Executive function.** The experimental participants in the Low EF group demonstrated significant training effects on cognitive performance trajectories that were of a large magnitude (Cohen's $d$ = 2.23). Whilst these participants were compared to a relatively low number of controls ($n$ = 8) and consequently interpreted with caution, they were nonetheless acceptable and meaningful (*e.g.*, Pietrzak et al., 2015). In addition to providing support for hypothesis one heterogeneity of cognitive performance, this result also provides support for hypothesis two – that a generalised effect of training would be identified. The EF results are in contrast to the present study's lack of training effects on memory (VM and LTVM), shown above with the VM task (measuring specific training effects), and the LTVM measure (also used to assess generalised effects).

The EF results are also supported by, and build on some of the training literature, albeit limited in the EF domain (Stuss et al., 2007; Winocur et al., 2007a; 2007b). For example, Stuss and colleagues (2007) combined memory strategy, goal management and self-efficacy in their training paradigm. This intervention led to improvements in *self-reported* EF. The present study therefore builds on this research showing *objective* evidence of EF performance increases. The results also extend from studies demonstrating positive objectively measured cognitive effects following multi-domain training in the distinct, yet similar setting of adult rehabilitation (Gehring et al., 2009; 2011). Gehring and colleagues (2009) reported the results of a RCT of cognitive rehabilitation in 140 adult patients with a glioma (cancerous glial cells of the brain; age *M* = 41.8 years, *SD* 9.5 years). They showed that the intervention group performed significantly better than the control group on a reliable change index (RCI) composite score of neuropsychological tests including EFs such as attention and inhibition, in addition to VM performance. These results were maintained for 6 months following the cognitive rehabilitation program. Thus the present study adds to the literature by more clearly demonstrating longitudinal effects of training on EF.

*Proportion of individuals gaining from EF training.* Interestingly, the Low EF class results are also comparable to the past literature in relation to the proportion of experimental participants demonstrating gains in other cognitive training studies (*e.g.*, Ball et al., 2002; Wu & Witkiewitz, 2008; Willis et al., 2006). The Low EF class in the present study consisted of *n* = 48 in total, representing 16.16% of the entire cohort. For example, as previously noted in the ACTIVE study only 26% of participants in the memory training group demonstrated significant improvement in subsequent memory testing (Ball et al., 2002; Rebok et al., 2014; Willis et al., 2006).

**Demonstration of Support for Training Effects Across Cognitive Domains**

It should also be highlighted that there is support for training efficacy overall (*i.e.*, across the VM, LTVM and LTVM outcome measures). The GGMM showed cognitive gains made by the experimental group from baseline to 12 months and positive slope trajectories (albeit non-significant for the VM and LTVM measures) in all the Low classes. That is, gains were made in those trained, not only in the Low performance classes but also in participants allocated to the High (and Moderate) performing classes whereby cognitive performance gains across VM, LTVM and EF were demonstrated. *Thus, importantly, this leads to the conclusion that the present study's limited evidence of memory benefit, and therefore lack of support for hypotheses one and two, may merely highlight limited success in revealing differences between experimental and control conditions*. Both statistical and recruitment issues may have contributed to the inability to discriminate between the conditions and are discussed in the limitations section below. As mentioned above, a lack of demonstrable gains compared to controls has been noted by others (*e.g.* Martin et al., 2011). Following their meta-analysis, Martin and colleagues (2011) conclude that training does lead to performance gains but none of the effects observed could be attributed specifically to cognitive training. They noted issues with the demonstration of improvements that exceed the improvement control conditions (Martin et al., 2011).

**Inter-individual Differences in Training Efficacy Across Cognitive Domains**

Whilst there was demonstration of support of training within all the outcome measures, the data also indicate differential treatment effects across measures. This is evidenced by differences in class enumeration for the VM measure and for the EF. This represents a unique finding, given there have been no other known specific studies noting the commonality (or lack thereof) of differences in treatment effects to memory and EF classes demonstrating training gains.

The results do, however support past indications of differential treatment effects on outcome measures across memory domains (*e.g.*, Fairchild et al., 2013; Langbaum et al., 2009; Project MATCH Research Group, 1999). In Langbaum, and colleagues' (2009) previously-noted study, one class was created in which the participants had a high conditional probability of responding to the HVLT, while participants in another class had highest conditional probabilities of responding for the RAVLT. The third class included individuals not responsive to training on either measure. The authors therefore concluded that there is variability in responsiveness to memory training across outcome measures, given they were able to distinguish three types of respondents.

Similarly, given the differences in class enumeration (*i.e*., there were more individuals allocated to the Low EF class demonstrating training gains than the Low VM class) it also indirectly suggests that different individuals must profit differently to training (Martin et al., 2011). Factors influencing these differences are further discussed when considering the results of the second GGMM incorporating baseline predictors as predictors of training trajectories, below.

**Low Performance at Baseline**

The findings that the experimental participants in the *Low* EF class demonstrated both statistically significant and large, clinically significant gains compared with controls are also supported in the literature (Fairchild et al., 2013; Langbaum et al., 2009; Lövdén et al., 2012; McKitrick et al., 1999). That is, those performing relatively lower at baseline demonstrate greater cognitive improvements following training than those higher-performing at baseline.

The previously mentioned study by McKitrick and colleagues (1999), training with community-dwelling older adults, found that participants with lower pre-training scores appeared to benefit from training programs compared with higher-performing older adults. Pretest performance was a predictor of gain across two outcome measures. Specifically, 63% and 61%, of trained individuals who were low-performing at baseline met the authors' success criterion for significantly improved performance levels post-intervention. In contrast, of those considered to be higher-performing at baseline, a much lower percentage of individuals (only 38% and 33% of individuals) indicated improvement following training.

More specifically relating to the longitudinal results of the current study, Fairchild and colleagues (2013) looked at the performance of those in McKitrick and colleagues' (1999) study over a 12-month period. In the Fairchild study, lower baseline performance was associated with greater improvements longer-term. The group that had the highest rates of success from training (67% of the cohort) had lower performance at baseline. The group that had the lowest rate of success (13%) had higher baseline performance.

Whilst the results for the Low classes across the three outcome measures are supported within the literature cited above, they appear counter to some more widespread evidence in the ageing literature (particularly in the context of mnemonic training) and dominant theories; specifically, that the benefits from training are often smaller for those who need training the most (Fandakova et al., 2012; Kliegl et al., 1990; Kramer & Willis, 2002; Lövdén et al., 2012; Stigsdotter-Neely & Bäckman, 1995; Verhaeghen et al., 1992). As noted in Chapter 4, Stigsdotter-Neely and Bäckman (1995) used hierarchical regression to look at multi-domain cognitive training of older adults ($n$ = 46), and found that *higher* pre-training scores (alone) reliably predicted the magnitude of training-related gains. Specifically, the pretest score explained 70% of the variance in the immediate post-training score for a task involving recall of concrete words and 56% of the variance at 6-month follow-up.

### Cognitive Training and the Compensation and Magnification Views

In addition to the present study's GGMM having support in the literature, the Low EF findings support the compensation view – a theory posited in the literature to conceptualise inter-individual differences in training gains (Lövdén et al., 2012). That said, close scrutiny of the findings of the present study in fact provide evidence to support the *simultaneous* presence of another theory posited to explain inter-individual differences in training gains – the magnification view (Lövdén et al., 2012). Whilst both the compensation and magnification views make distinct predictions, their concurrent presence is also considered to be possible because neither account includes predictions as to the conditions under which they may or may not apply. This gives room for *post-hoc* explanations of empirical observations (Lövdén et al., 2012; Baltes, 1987). Indeed the literature demonstrates support for both conceptualisations, as well as the possibility of their simultaneous presence (Lövdén et al., 2012).

The *compensation view* predicts that individuals who gain from cognitive training are those with lower initial performance and cognitive ability. That is, training compensates for their deficits. The corollary suggests that individuals with a good level of performance and ability initially do *not* improve following training. These higher-performing individuals already function at an optimal level and could be considered to be successfully ageing. That is, they exhibit a ceiling level of cognitive functioning for their age and therefore have less room for improvement (*e.g.*, Fairchild et al., 2013; Lövdén et al., 2012). The compensation view may be evidenced, for example, by a negative association between training gains and both cognitive abilities and initial levels of performance.

In contrast to the compensation view, the *magnification view* highlights increasing inter-individual differences in cognitive performance following cognitive training. Individuals who show higher performance on cognitive ability measures, and who demonstrate higher baseline performance in memory, gain more from cognitive training (Lövdén et al., 2012). This view suggests that individual and age-related differences in gains from cognitive training can be explained by initial differences in cognitive capacity. The difference in ability enables those individuals to acquire, implement and sharpen effortful cognitive strategies through training. Thus, for example, there is a positive association between cognitive gains and initial performance. The magnification view is more prominent in the literature when interpreting age differences after mnemonic training (Kliegl et al., 1990; Lövdén et al., 2012; Verhaeghen & Marcoen, 1996).

**Consistencies with the compensation view.** The present data based on baseline performance appear to largely support the compensation view (versus the magnification view) given that the trained individuals in the Low EF performance class showed statistically significant gains of performance in comparison to the controls, and that the difference between the slopes was of a large magnitude. Indeed, whilst it is not a specific criterion of the compensation view, those allocated to the Low EF class could also be considered low-performing from a normative perspective (a Low Average level; Strauss et al., 2006).

The data also support the compensation view for each of the three outcome measures (*i.e.*, VM, LTVM, and EF) in those individuals assigned to the High performing classes, when considering baseline performances.[53] That is, the trained individuals allocated to these classes had less room for improvement and therefore did not demonstrate statistically significant gains following intervention. This is despite these classes having an inadequate number of controls to which the experimental groups were compared for the VM and LTVM. For example, the GGMM for VM demonstrated the potential for ceiling effects. The High VM class recalled, on average, 13.6 words out of a possible 15 words at baseline. This represents 1 *SD* above the mean.[54] The same analysis for the LTVM outcome measure showed that the High LTVM class recalled 12.4 words at baseline, also 1 *SD* above the mean.

Finally, the High performance EF class was demonstrating baseline performance at an 'Average' level in comparison to norms (CogState Ltd, 2015). Thus it could be argued that these individuals were already successfully ageing prior to training given that, by definition, those who are successfully ageing show little or no decrement in cognitive functioning in domains typically seen in epidemiological, cross-sectional and longitudinal studies of decline (Barnes et al., 2007; Carey, 2007; Jones et al., 2005; Lindenberger & von Oertzen, 2006 in Raz, 2009; Nelson & Dannefer, 1992; Raz, 2009; Schaie, 1994; Yaffe et al., 2009).

Furthermore, the participants allocated to the High and Moderate VM and LTVM classes, and those in the High EF class were considered, on average, to be younger-old adults who had high levels of CR, as shown by their above-secondary levels of education and High Average estimated premorbid IQ. These individuals therefore

[53] Here the VM and LTVM growth mixture models incorporating training status as a predictor are interpreted given the overall acceptability of the models, and an adequate number of controls allocated to theses classes.

[54] The RAVLT has also been shown to have ceiling effects, including Trial 5 data (Graf & Uttl, 1995; Uttl, 2005). This will be further discussed in the limitations section.

would have high levels of inherent ability.[55] Here the Moderate and High performance classes' lack of improvement is likely due to the fact that initially, these individuals had capacity and demonstrated this cognitive capacity at baseline with higher cognitive performance (Salthouse, 2006).

**Consistencies with the magnification view.** As noted, there is also some support for aspects of the magnification view, in addition to features of the compensation view. Closer scrutiny of the baseline characteristics of those individuals allocated to the Low performance classes across the three outcome measures utilised in the current study shows that they consisted of younger-old adults,[56], and those with above-secondary education.[57]. A nuanced view suggests that these individuals had cognitive capacity/ability given their younger age and above secondary education. That is, they have intact cognitive resources that aid performance as others have suggested (Lövdén et al., 2012). Whilst the magnification view argues that those with high ability are able to utilise training and demonstrate *greater* gains than those *without* high ability, the current data supports the magnification view pattern of findings whereby the gains from cognitive training can at least in part, be explained by initial differences in cognitive resources (Lövdén et al., 2012).[58].

Similarly, according to CR theory, under active models, high reserve individuals should have more flexible brain structures, cognitive processes and/or knowledge structures to enable training benefits (Tucker-Drob & Salthouse, 2011). It is these

[55] Again this is further explored when discussing the second GGMM investigating baseline characteristics as predictors. This is discussed further below.

[56] Older adults can be defined as 'younger-old' (ages 65-74), 'older-old' (ages 75-84), and 'oldest old' (ages 85+; APA, 2009).

[57] Low EF class age: $M$ = 69.64 years. Low EF class estimated premorbid IQ: high Average $M$ = 109.4.

[58] As noted, further direct exploration of the impact of cognitive resources was conducted with the second GGMM conducted, incorporating baseline characteristics as predictors for cognitive trajectories. This is discussed further below.

high levels of reserve that are likely to assist them to acquire and implement training protocols and perhaps sharpen their use of the strategies introduced (*e.g.*, Baltes, 1987; Fandakova et al., 2012; Gilsky, Rubin, & Davidson, 2001; Lövdén et al., 2012). The data is similarly supportive of suggestions that high levels of CR equate to learning efficiency and thus demonstrate positive training effects compared with controls (Ferguson, 1956; Stern et al., 1994; Sullivan, 1964).

The results in this study are consistent with some of the more widespread evidence in the ageing literature with regard to training, which indicates that higher overall baseline abilities can predict training gains in healthy older adults[59] (Fandakova et al., 2012; Kliegl et al., 1990; Kramer & Willis, 2002; Lövdén et al., 2012; Stigsdotter-Neely & Bäckman, 1995; Verhaeghen et al., 1992). For example, Fandakova and colleagues (2012) demonstrated that older adults who showed more improvement following training also had a preserved ability to initiate strategies to facilitate performance.

Thus, as noted, key to interpreting these findings is that the data supports the *simultaneous presence of elements of both the magnification and compensation views*. In particular, those who are in the EF Low class at baseline demonstrated statistically significant improvements of a large magnitude following training, consistent with the compensation view. That is, their relative low baseline performance influences performance gains. The non-significant gains demonstrated by the trained individuals in the High classes in the VM, LTVM and EF growth models indicated they had less room for improvement due to a ceiling effect, also supporting the compensation view. In contrast the magnification view is supported due to the experimental group in the

[59] That is, the Low EF class had high capacity. The High class with even greater levels of CR may have experienced ceiling effects at baseline precluding the detection of improvements.

Low EF class having low baseline performance and high levels of pre-training ability enabling training gains.

**Consistencies with the 'Use It or Lose It' Hypothesis and Associated Theories**

With demonstrable improvements in cognitive trajectories following training across domains, and particularly with the acceptable EF GGMM showing generalised training effects, the present research findings also lend some support to the 'Use it or lose it' hypothesis and associated theories overall, including the disuse theory (Salthouse, 1991). The provision of a more cognitively enriching environment through training to largely 'underperforming' individuals (those in the Low performance classes) enables cognitive performance gains following training. Those in the Low performance classes also perhaps lacked a cognitively enriching environment to maintain their cognitive function, *i.e.*, they were experiencing disuse of their cognitive resources (Hertzog et al., 2009; Salthouse, 1991). In addition, the generalised growth mixture models' demonstration that training induces positive cognitive changes can be interpreted as evidence of plasticity and increased CR (Brehmer et al., 2008; Noack et al., 2009; Brehmer, et al., 2007; Lövdén et al., 2012). This will be specifically discussed in relation to hypothesis three below, where the effects of baseline characteristics on training effects were directly explored.

**Statistical Implications: Caution with Interpretation of Generalised Growth Mixture Modelling Data Class Labels**

Importantly, the present study also highlights further the need for researchers to be cautious when interpreting GGMM data. Specifically, conclusions about Low (Moderate) and High performance are drawn relative to the cohort (and labels assigned accordingly). Clearly this is the case for all studies, however it is particularly important to draw attention to this, given that the GGMM technique is relatively new. The GGMM technique forms groups of individuals with homogenous trajectories and makes them distinct from other homogenous trajectories within the *cohort*. Thus conceptualisations of the classes formed are particularly necessary when making links to past findings and theory. It is important to consider the value or level of the predictor beyond its relativity to the cohort.

Thus, whilst statistically it was appropriate to follow convention and label the classes Low, (Moderate) and High performance classes, it is important to consider the meaning of those categories in the context of past research and relevant theoretical models, and not only in optimal model selection (Andruff et al., 2009; Bauer, 2005; Jung & Wickrama, 2008; Nagin & Odgers, 2010) but also when drawing conclusions. This is particularly important in the training context if these models are to become more commonplace.

**Clinical Implications**

Despite the statistical caution required, clinically, investigation of longitudinal performance trajectories through GGMM techniques can be useful to inform facilitators and participants alike of the inter-individual temporal effects of training. That is, it provides preliminary information with regard to the expected pattern of change over time and ultimately clinical guidance for how best to optimise individualised training allocation to achieve executive performance gains (Gueorguieva et al., 2007; Kreuter & Muthén, 2007; Stulz et al., 2010). There is currently no known information of this kind to inform program facilitators and participants alike of inter-individual temporal effects of training.

### Summary of Results in Relation to Hypotheses 1 and 2

The current findings are a significant contribution to the cognitive training literature. Generalised growth mixture modelling has been highlighted as important when determining the nature of the efficacy of the ACE program, a multidomain cognitive training intervention.

Firstly, heterogeneity of longitudinal cognitive performance trajectories following training was demonstrated with EF performance trajectories. This supports past indications of the need to consider inter-individual responsiveness, and that some generalised effects of training are demonstrable. In the past, the effect of training on EF has been unclear due to a limited number of studies investigating this cognitive domain. The model provides a solid foundation on which further research can build.

Secondly, importantly, there was no meaningful evidence of specific longitudinal VM performance trajectory gains following training, or generalised effects of training on memory (LTVM). The efficacy of training on memory was inadequately recovered by the GGMM given that the control sample was not considered large enough to adequately extrapolate any gains to other populations.

Thirdly, whilst there is no evidence for meaningful performance gains in memory, there is evidence for the training efficacy overall across experimental participant VM, LTVM and EF. That is, there were positive slope trajectories for experimental participants across all classes – High (Moderate) and Low and across all cognitive domains VM, LTVM and EF. Thus, importantly, the present study's limited evidence of memory benefit may merely highlight limited success in revealing differences between experimental and control conditions.

The results provide some support of the ‘Use it or lose it’ hypothesis and associated theories with cognitive gains in those trained – particularly with meaningful executive performance gains. This can be interpreted as evidence of plasticity and increased CR. Further support for these theories came from the identification of the characteristics of individuals who demonstrated improvement, discussed below.[60]

Finally there are both statistical clinical implications of the study. The findings show that caution is required when interpreting GGMM class labels, which are relative to the cohort, when drawing theoretical conclusions. Clinically, the results are useful to inform facilitators and participants alike of the inter-individual temporal effects of training. This is particularly important if it is to become more commonplace in the ageing and training context.

**Hypothesis 3: Baseline Characteristics Predictive of Cognitive Trajectory Gains**

The second application of GGMM further explains the classes, beyond the relative labelling of classes and more general consideration of baseline characteristics. The analyses incorporated age, sex, and proxies for CR – education and estimated premorbid IQ as predictors of the inter-individual cognitive trajectories across the three outcome measures (VM, LTVM and EF). However, given the previously noted inadmissibility of the VM and LTVM models due to the low number of controls to which the experimental participants were compared, the discussion largely focuses on the EF results. The EF results support the third hypothesis of the study: that individual baseline characteristics were predictive of interindividual cognitive trajectories. Specifically the Low EF class was more likely to be older and have a lower estimated

[60] The second GGMM explored baseline characteristics as predictors of cognitive gains, discussed below.

premorbid IQ in comparison to the High EF class. Education and sex were not found to be predictors.

**Support for the Model Utilised**

Two distinct trajectories (High and Low performance classes) were identified for the EF outcome measures. Consistencies with class enumeration with the previous GGMM model (incorporating training status as a predictor) supported the validity of incorporating baseline characteristics as a predictor in the second GGMM model, given that finding the same number of classes across models was not guaranteed.[61]. It strengthens the substantive use of the GGMM technique when drawing theoretical justifications from the present study's data to past theory, such as CR, and the compensation and magnification views. With such substantiation, findings can also be used to provide further *a priori* clinical guidance regarding the expected patterns of change over time for particular individuals, and thus how best to optimise treatment selection (*e.g.*, Gueorguieva et al., 2007; Keller, 2001; Kreuter & Muthén, 2007; Lutz et al., 2006).

This study used the Low performance classes as the reference against which to compare the influence of predictors on performance trajectories. It is statistical convention to select one class with which to compare predictive effects (*e.g.*, Stulz et al., 2010). The Low class was chosen given that this was the class demonstrating statistically significant trajectory gains in the experimental group compared with control in the first GGMM.

[61] Thus, the reverse can also be stated, whereby the class enumeration in the second GGMM supported the class enumeration in the first GGMM conducted.

The EF findings represent a unique contribution to the ageing literature, given that no other study has examined inter-individual differences in EF performance trajectories. They will be further discussed below.

**Age**

As noted age was a predictor of performance supports the hypothesis of its predictive effect on the distinct cognitive performance trajectories. The EF model demonstrated that the individuals allocated to the Low EF class was significantly *older*[62] than those allocated to the High EF class. However, with this model, again caution is taken when interpreting the EF data and drawing conclusions given the age level is relative to the cohort. Once again, it is important to consider the value or level of the predictor beyond its relativity to the cohort when reaching conclusions. Whilst the individuals allocated to the EF class demonstrating cognitive gains were significantly older, according to the APA they would be considered "younger-old" adults (*i.e.*, 65–74 years; APA, 2009).[63].

With this nuanced interpretation, the results are consistent with the prominent view across previous research (Brown et al., 2003; Gross et al., 2012; Lovelace & Twohig, 1990). It supports studies with older adults of relative younger age being associated with more gains from training (Boron et al., 2007a; Sheikh, Hill, & Yesavage, 1986; Verhaeghen et al., 1992; Zelinski et al., 2008). Previous studies have also shown older age is cross-sectionally associated and predictive with lower initial performance (Jones et al., 2005; Langbaum et al., 2009).

[62] Low EF class age $M = 70.36$ years.

[63] As noted, older adults can be defined as younger-old (ages 65-74), older-old (ages 75-84), and oldest old (ages 85+; APA, 2009).

The data are consistent with the literature showing that individuals of a similar age are capable of plasticity, demonstrable by cognitive improvements (*e.g.*, Brown et al., 2003; Gross et al., 2012; Jessberger & Gage, 2008; Kempermann et al., 1998; Singer et al., 2003).[64] For example, Singer and colleagues (2003) demonstrated that those aged in their 60s and 70s showed significant training gains. Similarly, Gross and colleagues (2012), implementing multiple-group latent growth models, found *younger* age predicted better performance following training in their cohort, in terms of a slower decline from initial recall. Those trained in Gross' study could be classified as younger-older adults (*M* age = 73.5).

The results are also consistent with the dominant view in the literature that with older-old age plasticity decreases, leading to fewer training gains (Baltes et al., 2006a; Brown et al., 2003; Carey, 2007; Hertzog et al., 2008; Jessberger & Gage, 2008; Jones et al., 2006; Kempermann et al., 1998; Lustig et al., 2009; Noack et al., 2009; Schaie & Willis, 2010).

This leads to the obvious question of why the trained individuals allocated to the High classes across all outcome measures (and the Moderate classes for the VM and LTVM)[65] did not demonstrate significant improvements. They arguably had even greater capacity for plasticity with significantly *younger* age.[66] As highlighted with discussion above in relation to the compensation view, the key to the interpretation of the current data is that initially these individuals were already performing at a very

---

[64] Initial levels of capacity will be further discussed below when considering proxies for cognitive reserve, education and estimated premorbid IQ, and their impact on cognitive trajectories.
[65] Here the VM and LTVM models' Moderate and High performance classes are interpreted given an appopriate number of controls to which the experimental participants were compared in these classes.
[66] Moderate VM class age *M* = 64.74 years; High VM class age *M* = 68.06 years; Moderate LTVM class age *M* = 64.53 years; High LTVM class age *M* = 67.75 years; High EF class age *M* = 65.87 years).

high level. They were demonstrating successful ageing with at least average level performances in relation to norms at baseline (Barnes et al., 2007; Carey, 2007; Jones et al., 2005; Lindenberger & von Oertzen, 2006; Nelson & Dannefer, 1992; Raz, 2009; Schaie, 1994; Yaffe et al., 2009). Peak cognitive performance is a consequence of having pre-existing maximal levels of plasticity (Salthouse, 2006).

Thus, the results of the second GGMM highlight that the High (and Moderate) performance classes, demonstrate *the simultaneous presence of the compensation and magnification views*. The peak levels of pre-existing capacity in supports the compensation view because the higher classes exhibited ceiling performances, whereas in the Low EF class the low performance at baseline showed enabled training gains. As the magnification view suggests, the Low EF class had some degree of the plasticity required as evidenced by a younger-old age classification. Thus, to enable significant training gains, individuals must possess high cognitive capacity concurrent with low baseline performance. Such conclusions can be drawn given the appropriate use of GGMM.

**Inconsistencies with Past Literature**

There are some differences between the current study's results and recent studies exploring these baseline characteristics. Some studies have shown differences in the direction of the association or predictive value of the baseline factors, for example, younger age was predictive of gains *vs.* older age (Fairchild et al., 2013; Gross et al., 2012; Langbaum et al., 2009; Schaie & Willis, 1986, 1994). Differences between the findings of the present study and past research may be due to a number of different factors. As noted, there is insufficient investigation in cognitive domains beyond memory (Martin et al., 2011). Also, as discussed, there lack of consideration of interindividual differences and cognitive trajectories (*e.g.*, Brehmer et al., 2008; Calero & Navarro, 2007; Dahlin et al., 2008; Nyberg et al., 2003; Zelinski et al., 2008) thus a potential distortion of results (Salthouse, 1991).

**Caution with interpretation of growth models: Consideration of trajectories.** When appropriate statistics are implemented, some studies still reveal inconsistencies with the present study's findings, however (*e.g.* Langbaum et al., 2009). Consistent with the present study, Langbaum and colleagues (2009) showed that on face value individuals who demonstrated training responsiveness were older. Unlike the present study, however, Langbaum and colleagues' responsive class was 1.83 times more likely to fall into the 75- to 84-year-old age range, compared with the reference group (65- to 74-year-olds). That is, they would be considered older-old adults in comparison to the rest of the cohort unlike the present study's younger-older adults.

Discrepancies in the interpretation of results can be explained by close attention to the trajectories of the classes used to draw conclusions. Firstly, Langbaum measured training across a shorter follow-up period (*i.e.*, pre-post *vs.* 12 months for the current study). Predictors for trajectories may differ according to the duration of the follow-up of the measured effect. Age may have different predictive effects longitudinally compared to more immediate effects. Such discrepancies in the predictive value of factors based on duration of follow-up have been highlighted in past studies implementing GBGM techniques (Longabaugh, Wirtz, Zweben, & Stout, 2001; Wu & Witkiewitz, 2008). Secondly, the inconsistent results may also be due to the gradient of the trajectories produced. Langbaum (and the ACTIVE study as a whole) demonstrated that there were in fact *declines* in cognitive performance in two of their three classes (the class with the second highest level of responsiveness, the HVLT class, and those classified as "non responders" (Langbaum et al., 2009, p. 21) across time, whilst the current study showed cognitive *gains* in all classes. Thus comparisons with their reference class – those who responded best, demonstrating positive trajectory gains – were compared to two classes which declined. Factors that predict gains may therefore differ when comparing declines to gains. Indeed when Langbaum compared the two lower responding classes, there was no predictive effect of age. Yaffe and colleagues (2009) also suggest that predictors of those individuals who maintain cognitive function in older age may be different from those impacting individuals who decline.

Thus, it appears necessary to take into the nature of cognitive performance trajectories (*i.e.*, the timeframe and gains versus declines) when drawing conclusions about the effect of age. It also shows that whilst more sophisticated statistics considering trajectories remain in their infancy in the ageing literature, more research is required.

**Sex**

The lack of predictive effect of sex on EF performance does not support the hypothesis made that females would demonstrate greater gains following training. The present study's results do, however, make a significant further contribution to the literature regarding the presence (or lack thereof) predictive effects of sex, given there are very few studies adequately investigating EF trajectories following training (Herlitz et al., 1997).

Despite the lack of studies investigating EF trajectories, the finding that sex is not a predictor is also consistent with past indications of the impact of sex at least in the *memory* domain (*e.g.*, Gross et al., 2012). Gross and colleagues (2012) also found that sex was not a predictor of learning curves across time. Like the present study, they used growth models.

There are also some consistencies with other memory findings by Gross and colleagues, and others (*e.g.* Jones et al., 2005). These authors demonstrated that sex predicted *baseline performance*. Specifically, both studies demonstrated that males had lower initial learning levels than did females. Whilst it was not investigated whether sex was predictive of baseline performance, the present study also demonstrated that the Low EF performance class, those labelled as such due to low initial performance, was more likely to consist of males. The present study demonstrated that it was low performance individuals who were more likely to show cognitive trajectory gains. Comparisons with memory performance made here are done so with prudence, however, given that specific predictors may have differential effects on distinct cognitive domains (*e.g.*, Fairchild et al., 2013 Langbaum et al., 2009; Project MATCH Research Group, 1999). That is, the predictive value of sex may be somewhat task dependent (*e.g.*, Schaie, 1994; Schaie & Willis, 1986).[67] In fact, this may further explain the discrepancies of the Low EF results with past research investigating performance in the memory domain and supports the notion that there are different effects of sex on performance (Brehmer et al., 2008; Gross et al., 2012; Zelinski et al., 2008).

## Proxies for Cognitive Reserve: Education and Estimated Pre-morbid IQ

**Estimated premorbid IQ.** As noted, individuals in the Low EF class were also more likely to have a lower estimated premorbid IQ compared with the High performing class. These findings therefore appear, on face value, to be incongruent with hypothesis three and some past research examining these characteristics, which suggest higher levels of CR enable greater training gains (Bagwell & West, 2008; Chu

[67] The present study investigated this further with analyses comparing those individuals demonstrating improvements across the domains investigated (*i.e.*, VM, LTVM and EF) and will be discussed further below.

et al., 2007; Green et al., 2008b; Gehring et al., 2011 Hertzog et al., 2009; Hill et al., 1995; Hunt, 1978; Langbaum et al., 2009; McKitrick et al., 1999; West & Hastings, 2011).

Once again, caution is taken with the conceptualisation (*i.e.*, level) of this proxy for CR on the cognitive trajectories produced with the GGMM techniques.[68] Inspection of the level of CR in individuals in the Low EF class who demonstrated significant training gains reveals that whilst they had *lower* levels of reserve relative to the cohort, they can be considered to have a *higher* level of CR relative to norms. Specifically, the individuals allocated to the Low EF classes had a High Average estimated premorbid IQ and above-secondary education.[69]

Thus, reconsidering the present study's results in this light when making comparisons to past research and theory, the data supports hypothesis three and past conclusions that high levels of proxies of CR are associated with training gains (Brehmer et al., 2007; Carter, 2002; Fandakova et al., 2012; Yesavage et al., 1988). In addition, individuals with high levels of intelligence who are also low performing initially, appear to have a greater capacity to learn and implement training protocols to affect statistically significant and meaningful gains in cognition, as others have shown (Baltes et al., 2006a; Lövdén et al., 2012; O'Hara et al., 2007; Yesavage et al., 1988). Yesavage and colleagues (1988) investigated a measure of intelligence (verbal ability as measured on the WAIS vocabulary subscale score) in a small cognitive training study ($n$ = 40 trained older adults; age $M$ = 67 years). They found that those who had a high level of verbal ability (subscale score $M$ = 50.2 out of a maximum 66) benefited most from training. Interestingly their "older" adults also could be

[68] As previously discussed in relation to age as a predictor of performance outcomes.
[69] Low EF education $M$ = 13.19 years; Low EF estimated premorbid IQ $M$ = 109.56 (High Average when rounded).

considered younger-old, as is the case for the present study (APA, 2009). The present study's second GGMM supports indications in the literature that relatively higher intelligence overall aids treatment gains. Considering indices of intelligence as a predictor of EF performance in particular following training, the data are consistent with the few comparable studies (*e.g.*, Brehmer et al., 2007; Carter, 2002). Carter (2002) conducted a study of 93 younger adults (undergraduate students) that demonstrated the importance of high levels of CR in knowledge gained from training. A *high* general reasoning ability (considered to be a measure of psychometric *g*) was significantly related to knowledge gained following lecture-based training (*g M* = 82.9 out of a maximum 100).

The key to interpreting the current study's data demonstrating that the trained individuals in the High classes (and the Moderate classes for the VM and LTVM)[70] across all outcome measures did not demonstrate still greater improvements (given they had even higher premorbid IQ[71]) is that they were additionally functioning with high levels of pre-existing capacity. This is a similar argument to that given in the context of age. The High EF class also exhibited successful ageing with Average level performances in relation to performance at baseline (Barnes et al., 2007; Carey, 2007; Jones et al., 2005; Lindenberger & von Oertzen, 2006; Nelson & Dannefer, 1992; Raz, 2009; Schaie, 1994; Yaffe et al., 2009). Thus the High-class individuals may represent

[70] Again the VM and LTVM models' Moderate and High performance classes are interpreted given an appopriate number of controls to which the experimental participants were compared in these classes.
[71] Above-secondary education: Low VM class, *M* = 12.42 years; Moderate VM class *M* = 13.23 years; High VM class *M* = 14.4 years; Low LTVM education *M* = 12.81 years; Moderate LTVM class *M* = 13.38 years; High LTVM class *M* = 14.47 years; Low EF class *M* = 13.19 years; High EF class *M* = 14.01 years. High Average estimated premorbid IQ (compared with same age norms and when rounded to the nearest whole IQ value for some measures; Sattler & Dumont, 2004): Low VM class, *M* = 110.32; Moderate VM class *M* = 110.10; High VM class *M* = 113.39; Moderate LTVM *M* = 110 (*M* = 109.86); High LTVM *M* = 114.03; High EF *M* = 110 (*M* = 109.56).

a subsection of the population who are at their maximal performance potential overall, considering both baseline performance and baseline ability (Salthouse, 2006).[72] In contrast, Low classes were distinguished from other individuals in the High class, with arguably more reserve capacity, because they had more scope for improvement given normatively low performance at baseline.[73]

These data therefore again reveal, to an extent, the *simultaneous* presence of the compensation and magnification views. The compensation view is supported given that the Low performance classes demonstrated significant improvement, and the higher classes exhibited ceiling performance. Whilst the magnification view argues that those with high ability are able to utilise training methodology and demonstrate greater training gains than those without, the current data are consistent with the view that the gains from cognitive training can, in part, still be explained by initial differences in cognitive ability (Lövdén et al., 2012).

This nuanced conceptualisation of the data clearly supports a number of the ways in which CR has been described in the literature. It supports its use as a heuristic to explain training gains and the theory that CR can act as a measure of an individual's overall learning potential (Bagwell & West, 2008; Hill et al., 1995; West & Hastings, 2011). The findings also strengthen the argument that high levels of CR equate to learning efficiency (Ferguson, 1956; Stern et al., 1994; Sullivan, 1964). Under active models, high reserve individuals should have more cognitive flexibility, cognitive processes and/or knowledge structures (Brehmer et al., 2007; Carey, 2007; Hill et al.,

[72] It is also noted that the control group for these classes demonstrated significant improvement and thus may have accounted for a lack of significant improvement in memory trajectories by those trained. This will be further discussed in the Limitations section.
[73] As noted, the Low VM class was performing 1 *SD* below the normative mean of VM performance at baseline and Low EF class at Low Average level compared to norms (Strauss et al., 2006).

2000; Jones et al., 2006; Lövdén et al., 2010; Noack et al., 2009; Tucker-Drob & Salthouse, 2011). Brehmer and colleagues (2007) conducted a training study with 108 participants consisting of children, younger adults and an older adult group (65–78 years). The authors reported that older adults relied on their intact crystallised intelligence to overcome deficits in impaired memory functions. The learning was described as being "controlled and goal directed" (Noack et al., 2009, p. 437), thus indicative that these individuals executed cognitive flexibility. Furthermore, a number of past studies have shown that individuals with higher levels of CR and who are younger-old (APA, 2009) are most likely to implement memory strategies taught within training programs (*e.g.*, Fandakova et al., 2012; Gilsky et al., 2001). The Low classes in this study also have high levels of intelligence and were younger-old, and those trained may have been utilising their ability to implement training protocols. This more cautious, nuanced view of the models and data produced in the present study also further supports the 'Use it or lose it' hypothesis and associated theories. As previously noted in relation to hypothesis one and two, it could be said that the individuals allocated to the Low memory class were likely to be experiencing sub-optimal conditions and/or a subsequent disuse of their cognitive systems. Such disuse results in negative plasticity. This would explain the Low performance class individuals' relatively low cognitive performance at baseline, when it would have otherwise predicted good baseline performance (*e.g.*, Lövdén et al., 2012; Stern, 2002). By providing these lower-performing, high-reserve individuals with environmental stimulation – such as that derived from participation in the ACE training program – it offset the effects of disuse and the negative effects of possible unhealthy behaviours on EF, often seen in older adults (*e.g.*, Artaud et al., 2013; Hallett, 2001; Park et al., 2007; Kramer & Willis, 2003). That is, these individuals were able to use their existing resources and initiate plasticity mechanisms to learn and implement training protocols to compensate for this disuse. Park and colleagues (2007) reasoned that the effects of engagement would be greatest for older adults who had not previously been highly engaged. Further support for this argument comes from Summers' (2010) finding that an increase in the use healthy lifestyle activities across the 12-month follow-up may have contributed to observed improvements in cognitive performance. That is, it may be that the Low performing individuals in the present study's cohort had not maintained an engaged lifestyle through a variety of avenues known to promote more successful cognitive ageing, but increased these behaviours following training (Hertzog et al., 2009). Further research into the present study's cohorts' pre-existing levels of healthy training activities may further substantiate these claims.

**Education.** The finding that estimated premorbid IQ, but not education, was a predictor of Low EF performance classes seems counter to theoretical expectations and the hypothesis made that higher education would predict greater training gains. As both measures are considered proxies for CR, one might expect that if one were found to be a predictor, the other would also be predictive of the performance trajectories. However, this disassociation between the two proxies has been previously observed in the training literature (*e.g.*, Gehring et al., 2011; Green et al., 2008b) and there has been some discussion that reserve is not a unitary construct (Stern, 2002). Instead, the differential effects of education and estimated premorbid IQ on EF performance in the present study demonstrates that CR appears to be multi-faceted. There may be a difference between inherent ability (intelligence) and environmentally derived CR from education. The contrasting finding here between education and intelligence as predictors of cognitive gains may also suggest that estimated premorbid IQ is a more useful predictor (*cf.* Jones et al., 2011); however, full discussion of these topics is outside the objectives of the current study.

**Summary of Results in Relation to Hypothesis 3**

Overall, the present study demonstrates that use of GGMM can explain and predict heterogeneity in cognitive performance trajectories using baseline characteristics. Specifically, EF performance gains can be predicted by age, and estimated pre-morbid IQ. GGMM can therefore be used to begin to interpret inter-individual responsiveness to cognitive training. .

The current study, therefore, fills a gap in the literature, given that there are no other studies investigating age, sex and proxies for CR as predictors of inter-individual differences in EF performance trajectories. With such substantiation, findings can also be used to provide further *a priori* clinical guidance regarding the expected patterns of change over time for particular individuals, and thus how best to optimise treatment selection (*e.g.*, Gueorguieva et al., 2007; Keller, 2001; Kreuter & Muthén, 2007; Lutz et al., 2006).

The findings show that the Low memory classes at baseline, consisted of a small number of trained individuals, who were more likely to be older and have lower levels of estimated premorbid IQ than the High classes.

A cautious, nuanced view of the GGMM data shows that individuals allocated to the Low EF class demonstrating cognitive gains and who according to the APA would be considered “younger-old” adults (*i.e.*, 65–74 years; APA, 2009), were also considered to have ‘High Average’ estimated premorbid IQ compared to norms. Thus, reconsidering the present study’s results in this light, the data are congruent with past research and theory. It also highlights that researchers need to be cautious when interpreting GGMM data

The individuals in the High performance class may represent a subsection of the population who are at their maximal performance potential overall, with 'average' performance. Those in the Low performance class were distinguished from individuals in the High performance class, with arguably more reserve capacity, because they had more scope for improvement given normatively low performance at baseline. This distinction, therefore, supports the simultaneous presence of both the magnification view and the compensation views. That is, initial ability supports cognitive gains following training (*i.e.*, High Average estimated premorbid IQ) together with low baseline performance, respectively. These findings also support the 'Use it or lose it' hypothesis, plasticity, flexibility and the disuse perspectives. It suggests that those trained utilised their capacity for plasticity and flexibility and could therefore learn and implement training protocols as evidenced by EF performance gains. Their capacity may not have been evidenced given low baseline performance.

Overall, the results suggest that inter-individual differences in training responsiveness can be best defined by levels of existing ability and baseline performance. That is, those older adults most likely to exhibit cognitive training gains are individuals with high levels of existing capacity, yet who are currently underperforming in EF performance.

## Limitations

The current study revealed the nature of the efficacy of the ACE cognitive training paradigm, with clear training effects on EF. The study also demonstrated the utility of baseline characteristic predictors used to ascertain the nature of heterogeneous training responsiveness. Nonetheless, the findings and conclusions drawn have some limitations. Whilst the issues regarding the controls to which experimental participants have been compared have been highlighted earlier,[74] further limitations are further discussed below.

**Statistical Methods Undertaken**

Generalised growth mixture modelling, a GBGM technique, was considered the statistic of choice to address limitations of conventionally used statistics, by taking both a variable and person-centred approach. As a relatively new statistical technique, however, the information on factors that might lead to spurious conclusions is not fully understood. The present study also demonstrated some of the limitations that are often known to occur, in terms of both finding and creating the optimal model, and issues with model non-convergence and local solutions. These issues related to the creation of optimal models, the number of participants in the Low (reference) classes, missing data and ceiling effects.[75]

**Creating the optimal model.** Firstly, there was a linear relationship assumed within the models. Using more complex, quadratic models to estimate trajectories would have been desirable; however a larger sample (and higher proportion of participants per class) would be required to obtain trustworthy parameter estimates (Stulz et al., 2010). Secondly, there were some non-convergence issues with increased

[74] The Low VM and LTVM classes had only 2 control participants. Training gains in these classes were therefore inadequately recovered by the GGMM. The Low EF class with 8 controls was considered to be of a small size, but the GGMM was deemed to acceptable and could be used to draw cautious conclusions about treatment effects. It was also found that controls demonstrated positive trajectories, indicating improvement in cognitive function.

[75] This is in addition to the previously noted caution required when interpreting GGMM data when class labels are drawn relative to the cohort.

model complexity. Specifically, the differences in education between the experimental and control groups could not be accounted for in the final GGMM, due to non-convergence.[76] This is common with GGMM (Qureshi & Fang, 2011). Recommended solutions for non-convergence were executed, such as increasing starts and holding residual variances at zero; however the model remained inadmissible (Muthén & Muthén, 1998–2010; Petras & Masyn, 2010). Finally, and perhaps most importantly, limitations were evident in the inadequacy of model fit indices as a guide to model selection. Despite following statistical convention suggesting admissibility of the GGMM across all cognitive domains, the low number of controls in the Low memory classes rendered the GGMM as inadequate in demonstrating training gains. In fact, as noted in Chapter 5, Nagin and Odgers (2010) highlight that there can be uncertainty regarding model applicability when optimal model selection is based purely on these model fit indices. Thus, in the present study training effects *vs.* controls could not be well recovered by the GGMM technique.

Nonetheless, despite the limitations of the utilising GGMM to assess training, it was appropriate overall for a number of reasons. Firstly, GGMM more appropriately addresses emerging evidence on inter-individual training responsiveness to cognitive training than conventional statistics (Deary et al., 2009a; Duncan et al., 2002; Hedden & Gabrieli, 2005; Lindenberger & von Oertzen, 2006; Lövdén et al., 2010; Park et al., 2007; Raz, 2009; Terrera et al., 2010; Yaffe, 2009). Generalised growth mixture modelling, particularly the EF analyses of training effects, and models identifying baseline characteristics as predictors of class membership provides a more valid indication of longitudinal inter-individual performance trajectory differences. It has been suggested that conventional statistical methods may obscure information

[76] The possible influence of this on the overall results is discussed further below.

regarding the heterogeneity of individual performances within a sample exhibiting robust change (Connell & Frye, 2006; Langbaum et al., 2009; Willis & Schaie, 1987). Generalised growth mixture modelling is also pertinent because it considers the different classes as representing meaningful strata or data that is sufficiently meaningful to be used to guide development of 'Use it or lose it' and associated theories (Gueorguieva et al., 2007; Kreuter & Muthén, 2007). The selection of the GGMM approach was further supported given the consistency with class enumeration across the LCGA and GGMM techniques (Muthén, 2005; Qureshi & Fang, 2011).

**High Capacity, High Performance Older Adults**

The recruitment of the relatively young, high capacity cohort as a whole is also a limitation of the present study. The cohort consisted of younger-old adults with, on average, tertiary education and High Average estimated premorbid IQ. This may have contributed to the findings. For example, a sample that includes participants with largely high education at baseline increases the likelihood of higher initial test performance (Fairchild et al., 2013). Indeed most of the sample was allocated to the High and Moderate VM and LTVM classes, and the High EF class.

Similarly, the limited number of individuals considered to have lower cognitive ability impacts on the extent to which the present study's data supports the magnification view. It seems likely that if the current study had been able to recruit individuals with a wider range of demographics, specifically older individuals with lower levels of CR, more individuals would have been allocated to the Low performance classes and firmer conclusions drawn with regard to the magnification view.

The recruitment problem, however, is not exclusive to the current study, and is in fact extremely common within the training literature. The samples recruited in a number of previous studies also included individuals with relatively homogeneous demographic and baseline cognitive characteristics as well as participants with generally higher capacity, thereby making examination of training response variability difficult (*e.g.*, Fairchild et al., 2013; Gehring et al., 2011; Gross et al., 2012; Langbaum et al., 2009; Martin et al., 2011; McKitrick et al., 1999; Rebok et al., 2007). Much of the ageing research in general includes participants with an average of > 10 years educational attainment (*e.g.*, Bosma et al., 2003; McKitrick et al., 1999; Tucker-Drob & Salthouse, 2009). This bias in recruited samples may be due to the fact that higher-functioning older adults are most able to attend multiple training and follow-up testing sessions over an extended period of time, for example, given association with better health of participants (Fairchild et al., 2013; Gross et al., 2012; O'Hara et al., 2007; Park et al., 2007). Other authors note that it is difficult to separate out the effect of selection, selective attrition and causal directionality, whereby individuals who may already have greater cognitive ability are more likely to be attracted to, and engage with, training (Gold et al., 1995; Hultsch et al., 1999; Schooler & Mulatu, 2001). Future research should aim to recruit individuals of a greater demographic spread to address this issue.

Nonetheless, the cognitive training gains shown, further supports past studies showing heterogeneity of training (*e.g.*, Baltes & Kliegl, 1992; Schaie et al., 1987; Willis & Nesselroade, 1990). It is, therefore, imperative that inter-individual differences in cognitive performance responsiveness are considered, and the appropriate statistical analyses utilised. Such analyses were used in the present study. Indeed the findings highlight the validity of the use of GGMM in the exploration of cognitive training effects.

**Improvements by Controls in All Classes**

As noted earlier, a clear limitation of the current study is that there were positive cognitive trajectories expressed by the controls in *all* classes. That is, positive trajectories were demonstrated by the High and Moderate VM and LTVM classes, and the High EF class in the final models. The separate LCGA of the experimental and control groups for each outcome measure also showed that the control groups demonstrated gains in their performance trajectories. Whilst it was deemed that for the VM and LTVM the comparisons with $n$ = 2 controls impeded the meaningfulness of the results, improvements by controls showing positive trajectories could also explain why these models all demonstrated non-significant gains in the experimental participants following training compared with controls. The phenomenon whereby controls demonstrate cognitive improvement is consistent with past intervention evaluations as noted earlier (Ball et al., 2002; Langbaum et al., 2009; Martin et al., 2011; Owen et al., 2010). For example, a meta-analysis that examined the effects of training on immediate and delayed memory recall, found that there were no significant differences between trained and no contact controls in five of the seven studies analysed (Martin et al., 2011). The ACTIVE study also reported that the control group demonstrated significant pre-post-test gains (Ball et al., 2002).

One explanation for the improvement in controls is *practice effects.* Longitudinal studies are particularly subject to the influences of practice effects – a type of carryover effect whereby the repeated administration of a test influences the results of subsequent testing sessions. That is, greater exposure to tasks can lead to improvements in task performance. Both the RAVLT and the CogState Ltd GMLT tasks used to measure memory and EF in the present study are specifically designed to minimise practice effects by, for example, the use of alternative task forms (CogState Ltd, 2015; Strauss et al., 2006). However, practice effects have been reported in studies utilising both memory and EF outcome measures (*e.g.*, Salthouse, 2012; Schaie & Willis, 1986).

Additional contributions to the improvements seen in controls, which may apply to both the RAVLT and the CogState Ltd GMLT, could be the phenomenon of 'imitation of treatment' by controls, in which individuals assigned to the control group spontaneously behave as if they were in the intervention group (Hertzog et al., 2008; Shadish, Cook, & Campbell, 2002). Furthermore, changes in mood, level of motivation (see Fandakova et al., 2012), issues of social engagement through interaction with staff, including a desire to please, can all lead to temporary improvements in performance (Green & Bavelier, 2003; Lövdén et al., 2010; Park et al., 2007; Zelinski et al., 2011).

Thus, overall, these effects may have contributed to the non-significant differences between controls and ACE participants in the Moderate and High performing classes of the current study. This may have precluded demonstration of additional gain in the experimental group as a result of training. Whilst it can be argued that the same levels of practice may be demonstrated by the experimental group, the greater performance gains exhibited by the ACE participants across 12 months, as shown by the larger effect sizes for these participants, suggests some training effect.

As noted, the controls in this study had a significantly greater number of years of education, which could not be accounted for in the final GGMM due to non-convergence. Participants with greater levels of education are more likely to exhibit improved cognitive functioning over time, regardless of whether they participate in a cognitive training program (Beatty, Mold, & Gontkovsky, 2003; Gehring et al., 2011; Gontkovsky, Mold, & Beatty, 2002). For example, in the Gehring and colleagues' study of the training effects in adults with glioma, participants' level of education was shown to be a general predictor of improvement in neuropsychological functioning, irrespective of whether patients undertook training (Gehring et al., 2011). It has also been shown that the ability to self-generate strategies is associated with higher education and 'younger-old' age (Saczynski, Rebok, Whitfield, & Plude, 2007). Indeed, those in the present study who improved consisted of 'younger-old' individuals with above secondary education, thereby equipping them with overall greater cognitive ability, flexibility and a greater capacity to learn and implement self-generated strategies. Thus it can be considered that the current study's controls were more likely to show improvement over time after multiple neuropsychological test assessments (Gehring et al., 2011).

Ironically, the demonstration of improvement in the controls' performance trajectories in the High and Moderate VM and LTVM classes, and in the High EF class could be argued to support the overall concept of 'Use it or lose it'. Specifically, with *any* active cognitive engagement (*e.g.,* participation in a research project), individuals can attain improvements in memory and EF, cognitive domains which have been shown to decline in both cross-sectional and longitudinal studies (Buckner, 2004; Carey, 2007; Deary, 2009; Hedden & Gabrieli, 2004; Kramer & Willis, 2002; Schaie, 2000; Thompson & Foth, 2005).

## Missing Data

A common issue in longitudinal research is how to deal with *missing data.* The present study was also subject to this limitation. However, missing data on observed variables was handled using FIML, which considers missing data as MAR.[77] This is widely accepted as a pragmatic method of handling missing data, is commonly implemented in GBGM and was appropriate in the context of the present study (Feldman et al., 1999; Nagin & Odgers, 2010). Importantly, as previously noted in Chapter 5, MAR is considered to hold even in situations where missing values have been imposed by the researcher as a part of the study design. This is the case in the present study (Graham et al., 2001). Whilst covariates with missing values substantively affect the results because that participant will be dropped from the estimation (Huang et al., 2010), there was only 0.06% of missing data for the estimated premorbid IQ measures, and no missing data for treatment status, age, sex or education. The results are therefore unlikely to have been substantially affected.

## Ceiling Effects for the Verbal Memory Outcome Measure

Limitations also arose from the selection of Trial 5 from the RAVLT to measure VM performance. As previously noted, the High VM class demonstrated very high levels of VM performance even at baseline, thus there was little room for improvement. Whilst the Trial 5 outcome measure has been used in past studies to indicate verbal episodic memory performance, and is usually found to be the best learning trial, as well as being considered as one of the most reliable outcome measures of the RAVLT (*e.g.*, Uchiyama et al., 1995; Vakil et al., 2010), there have been reports it is affected by ceiling effects in healthy adults (Graf & Uttl, 1995; Uttl, 2005). Given that overall

[77] Muthén, Jo and Brown (2003) discuss nonignorable missing data modelling using missing data indicators.

the cohort in the current study consisted of younger-old, high-capacity individuals, it could be argued that the participants were also subject to ceiling effects for this outcome measure. These ceiling effects may have impeded accurate measurement of any cognitive training gains, as noted with regard to the compensation view.

Similarly, Magalhães and Hamdan (2010) demonstrated that there are significant effects of age and years of education on all RAVLT measurements (with the exception of the recognition trial). Fairchild and colleagues (2013) also highlighted that their sample consisted of younger-old adults (age $M$ = 64.7 years) and highly educated participants (education $M$ = 16.5 years). The authors suggested that these demographics may have contributed to ceiling effects, and suggested that a different pattern of results may emerge in an older or less educated sample. Indeed, the same could be said of the current study's results. Future research should consider using outcome measures with higher ceilings (and perhaps of greater difficulty) to account for high-functioning cohorts who demonstrate high baseline performance.

The consistency of the class enumeration of the VM and LTVM, however, reduces the extent of this problem in the current study. The LTVM measure has a higher ceiling than the VM measure and therefore enables greater 'room to improve'. With the same number of classes, and comparable number of experimental participants allocated to the Low performance classes across the outcome measures, the extent of this problem is reduced.

Overall, despite the limitations, any misinterpretation of conclusions made with regard to training effects and their subsequent application to guide intervention (*e.g.*, for training allocation) is unlikely to cause harm (*i.e.*, fulfilling the ethical psychological requirement to do no harm). This is particularly the case given the cognitive benefits from training seen across the classes and treatment groups for each of the outcome measures.

## Conclusion

The current study makes a significant contribution to the cognitive training literature, despite the limitations discussed above. The study used GGMM to identify and predict inter-individual differences in longitudinal cognitive performance trajectories in older adults following training. Generalised growth mixture modelling is a relatively new statistical technique that has fundamentally altered how we conceptualise and study change across time. The current research has shown the applicability and necessity of GGMM in the training context. Specifically, the use of GGMM in this study has led to an enhanced understanding of the efficacy of cognitive training on longitudinal performance trajectories and for whom training effects take place.

Importantly, the present study, using more appropriate statistics than past research, confirmed indications in the literature that there are distinct, inter-individual patterns of longitudinal cognitive performance following cognitive intervention.

The growth modelling revealed demonstrable inter-individual differences in training gains by experimental participants, including those of a significant and large magnitude for EF performance trajectories. This occurred for individuals allocated to the Low performance class, that is, those performing at a low normative level at baseline. These results therefore offer a novel contribution to the literature, given EF performance trajectories have not previously been explored.

The modelling also showed that a small number of individuals, those allocated to the relatively and normatively Low performing classes, demonstrate gains in VM and LTVM performance trajectories. However, the growth models incorporated only small numbers of controls to which experimental participants were compared. Thus no meaningful training effects on memory were demonstrated. The GGMM models therefore showed that the multidomain ACE cognitive training program produced some generalised cognitive improvement in healthy older adults, albeit to a limited extent.

The present study also identified that age and estimated premorbid IQ (proxies for CR) are predictive of inter-individual differences in longitudinal EF performance trajectories. These findings further highlight the utility of GGMM as a statistical method, beyond conventional statistics, for not only identifying, but also predicting these heterogeneous cognitive performances. This is particularly warranted in the cognitive training context given the need for further understanding of inter-individual responsiveness to training.

Importantly, from a statistical perspective, the present study underscores that when utilising GGMM, caution must be applied to interpretation of a number of factors. As noted, consideration of the number of controls when drawing conclusions is necessary beyond apparent statistical acceptability based on statistical convention. Consideration of class labels as well as the value or level of predictors is also required. Specifically, labels and the predictive factors identified are relative to the cohort. This cautious approach, in addition to attention to the nature of the trajectories produced (*i.e.*, the timeframe measured and gradient), is essential when comparing results to past research and theory. This is particularly necessary in the training context, in which these modelling techniques are beginning to become more commonplace and when such modelling is used to guide practice.

Whilst further corroborative quality research is necessary, the present study's findings utilising GGMM provide a solid basis for demonstrating inter-individual differences in training responsiveness and classification of individuals who significantly benefit from a training intervention, such as the ACE program. The results are therefore statistically useful.

Overall the study also reveals support for a number of prominent neurobiological and neuropsychological theories in the cognitive training literature. The results suggest that inter-individual differences in training responsiveness can be best defined by levels of existing ability and baseline performance. Specifically, the study indicates support for the simultaneous presence of elements of both the magnification view and the compensation views. Inherent ability (*i.e.*, younger-old age and High Average estimated premorbid IQ) predicts cognitive gains following training together with relative underperformance in a particular cognitive domain (*i.e.*, low baseline performance), respectively. That is, the older adults most likely to exhibit cognitive training gains are those individuals with high levels of existing capacity, and who are not demonstrating maximal performance prior to training.

These findings also support the 'Use it or lose it' hypothesis, plasticity, flexibility and the disuse perspectives that dominate the ageing literature. The provision of a more cognitively enriching environment through training to individuals with capacity for plasticity and flexibility, enabled individuals to learn and implement training protocols and demonstrate cognitive performance gains. Their capacity may not have been fulfilled given low baseline performance. The study is therefore conceptually useful.

Finally, the results offer clinically utility if inter-individual temporal effects of training are shared with training facilitators and participants alike. The findings can also be used in clinical decision making when allocating individuals to training programs and enhance cost-effectiveness of programs. Such critical information has been lacking to date.

# REFERENCES


Access Economics (2001). *Population Ageing & the Economy*. Canberra, ACT: Commonwealth Department of Health and Ageing.

Access Economics (2004). *Delaying the Onset of Alzheimer's Disease: Projections and Issues*, Report for Alzheimer's Australia, Canberra.

Acevedo, A., & Loewenstein, D. A. (2007). Nonpharmacological cognitive interventions in ageing and dementia. *Journal of Geriatric Psychiatry & Neurology*, *20*(4), 239–249. doi:10.1177/0891988707308808

Acock, A. C. (2005). *Growth Curves & Extensions Using MPlus*. Lecture Notes for Presentation at Academia Sinica, Taipei, Taiwan. http://people.oregonstate.edu/~acock/growth-curves

Agrigoroaei, S., & Lachman, M. E. (2011). Cognitive functioning in midlife and old age: Combined effects of psychosocial and behavioral factors. *Journals of Gerontology: Psychological Sciences.* doi:10.1093/geronb/gbr017

Albert, M. S., & Teresi, J. A. (1999). Reading ability, education, and cognitive status assessment among older adults in Harlem, New York City. *American Journal of Public Health*, *89*(1), 95–97.

Alexander, G.E., Furey, M.L., Grady, C.L., Pietrini, P., Mentis, M.J., Schapiro, M.B. (1997). Association of premorbid function with cerebral metabolism in Alzheimer's disease: Implications for the reserve hypothesis. *American Journal of Psychiatry*, *154*, 165–172.

American Psychiatric Association. (2000). *Diagnostic & Statistical Manual* (4th ed, text revision). Washington, DC: American Psychiatric Association.

American Psychological Association (2009). *Why Practitioners Need Information About Working With Older Adults.* www.apa.org/pi/ageing/practitioners/why.html

Anderson, I., Adinolfi, C., Doctrow, S., Huffman, K., Joy, K. A., Malfroy, B., ... Barnes, J. C. (2001). Oxidative signalling and inflammatory pathways in Alzheimer's disease. In *Biochemical Society Symposia* (*67*, pp. 141–150). London, UK: The Biochemical Society.

Andruff, H., Carraro, N., Thompson, A., Gaudreau, P., & Louve, B. (2009). Latent class growth curve modelling: A tutorial. *Tutorials in Quantitative Methods for Psychology 2009, 5*(1), 11–24.

Anstey, K., & Christensen, H. (2000). Education, activity, health, blood pressure and apolipoprotein E as predictors of cognitive change in old age: A review. *Gerontology, 46*(3), 163–77.

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.). *Advanced Structural Equation Modeling: Issues & Techniques* (pp. 243–277). Mahwah, NJ: Lawrence Erlbaum.

Arbuckle, T. Y., Gold, D., & Andres, D. (1986). Cognitive functioning of older people in relation to social and personality factors. *Psychology & Aging,* 1, 55–62. doi:10.1037/0882-7974.1.1.55

Artaud, F., Dugravot, A., Sabia, S., Singh-Manoux, A., Tzourio, C., & Elbaz, A. (2013). Unhealthy behaviours and disability in older adults: Three-city Dijon cohort study. *British Medical Journal*, 347:f4240. doi:10.1136/bmj.f4240

Attix, D. K., Story, T. J., Chelune, G. J., Ball, J. D., Stutts, M. L., Hart, R. P., & Barth, J. T. (2008). The prediction of change: Normative neuropsychological trajectories. *Clinical Neuropsychologist*, *23*(1), 21–38, doi:10.1080/ 13854040801945078

Australian Bureau of Statistics. (2008). *Population Projections, Australia, 2006 to 2101*. www.abs.gov.au/Ausstats/abs@.nsf/mf/3222.0

Australian Bureau of Statistics. (2009). *Australian Social Trends: Future Population Growth & Ageing.* Canberra, ACT: ABS.

Bagwell, D. K., & West, R. L. (2008). Assessing compliance: Active versus inactive trainees in a memory intervention. *Clinical Interventions in Aging, 3*(2), 371–382.

Baldi, R. A., Plude, D. J., & Schwartz, L. K. (1996). New technologies for memory training with older adults. *Cognitive Technology, 1,* 25–35.

Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., ... Willis, S. L. (2002). Effects of cognitive training interventions with older adults. *Journal of American Medical Association*, *288*(18), 2271–2281. doi:10.1001/jama.288.18.2271

Ball, K., Edwards J.D., Ross L.A., McGwin G., Jr. (2010). Cognitive training decreases motor vehicle collision involvement of older drivers. *Journal of the American Geriatrics Society. 58*(11), 2107–2113.

Baltes, P. B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. *Developmental Psychology*, *23*(5), 611–626.

Baltes, P. B., Freund, A. M., & Li, S-C. (2005). The psychological science of human ageing. In M. L. Johnson, V. L. Bengtson, P. G. Coleman, & T. B. Kirkwood (Eds.), *The Cambridge Handbook of Age & Ageing* (pp. 47–71). New York, NY: Cambridge University Press.

Baltes, P. B., & Kliegl, R. (1992). Further testing of limits of cognitive plasticity: Negative age differences in a mnemonic skill are robust. *Developmental Psychology, 28*(1), 121–125. doi:10.1037/0012–1649.28.1.121

Baltes, P. B., & Kunzmann, U. (2003). Wisdom. *The Psychologist, 16(3),* 131–132.

Baltes, P. B., & Lindenberger, U. (1988). On the range of cognitive plasticity in old age as a function of experience: 15 years of intervention research. *Behavior Therapy, 19*(3), 283–300.

Baltes, P. B., Lindenberger, U., & Staudinger, U. M. (2006a). Life span theory in developmental psychology. In R. M. Lerner (Ed.), *Handbook of Child Psychology: Theoretical Models of Human Development* (6th ed., Vol. 1, pp. 569–664). New York: Wiley.

Baltes, P. B., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. R. Nesselroade, & P. B. Baltes (Eds.), *Longitudinal Research in the Study of Behavior & Development* (pp. 1–40). New York, NY: Academic Press.

Baltes, P. B., Reuter-Lorenz, P., & Rösler, F. (Eds.). (2006b). *Lifespan Development and the Brain: The Perspective of Biocultural Co-Constructivism.* Cambridge, UK: Cambridge University Press.

Baltes, P. B., Staudinger, U. M., & Lindenberger, U. (1999). Lifespan psychology: Theory and application to intellectual functioning. *Annual Review of Psychology*, *50*(1), 471–507.

Barnes, D. E., Cauley, J. A., Lui, L. Y., Fink, H. A., McCulloch, C., Stone, K. L., & Yaffe, K. (2007). Women who maintain optimal cognitive function into old age. *Journal of American Geriatrics Society, 55*(2), 259–264. doi:10.1111/j.1532–5415.2007.01040.x

Barnes, L. L., Mendes de Leon, C. F., Wilson, R. S., Bienias, J. L., & Evans, D. A. (2004). Social resources and cognitive decline in a population of older African Americans and Whites. *Neurology, 63*(12), 2322–2326. doi:10.1212/01. WNL.0000147473.04043.B3

Barulli D., & Stern, Y. (2013). Efficiency, capacity, compensation, maintenance, plasticity: Emerging concepts in cognitive reserve. *Trends in Cognitive Sciences, 17*(10), 502-509. doi:10.1016/j.tics.2013.08.012

Basak, C., Boot, W. R., Voss, M. W., & Kramer, A. F. (2008). Can training in a real-time strategy videogame attenuate cognitive decline in older adults? *Psychology & Aging, 23,* 765–777.

Bauer, D. J. (2005). A semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling, 12,* 513–535.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods, 8,* 338–363.

Beatty, W. W., Mold, J. W., & Gontkovsky, S. T. (2003). The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) performance: Influences of sex and education. *Journal of Clinical & Experimental Neuropsychology, 25*(8), 1065–1069.

Belleville, S. (2008). Cognitive training for persons with mild cognitive impairment. *International Psychogeriatrics*, *20*, 57–66. doi:10.1017/S104161020700631X

Belleville, S., Gilbert, B., Fontaine, F., Gagnon, L., Ménard, E., & Gauthier, S. (2006). Improvement of episodic memory in persons with mild cognitive impairment and healthy older adults: Evidence from a cognitive intervention program.

*Dementia & Geriatric Cogntive Disorders*, *22*(5–6), 486–499. doi:10.1159/000096316

Bhardwaj, R. D., Curtis, M. A., Spalding, K. L., Buchholz, B. A., Fink, D., Bjork-Eriksson, T., Frise´n, J. (2006). Neocortical neurogenesis in humans is restricted to development. *Proceedings of the National Academy of Sciences of the United States of America*, *103,* 12564–12568.

Bherer, L., Kramer, A. F., Peterson, M. S., Colcombe, S., Erickson, K., & Becic, E. (2005). Training effects on dual-task performance: Are there age-related differences in plasticity of attentional control? *Psychology & Aging, 20*(4), 695–709.

Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, ageing, and cognitive control: Evidence from the Simon task. *Psychology & Aging, 19*(2), 290–303.

Bissig, D., & Lustig, C. (2007). Who benefits from memory training? *Psychological Science, 18*(8), 720–726.

Black, S.A., & Rush, R.D. (2002). Cognitive and functional decline in adults aged 75 and older. *Journal of the America Geriatrics Society*. *50*, 1978–1986.

Blackburn, J. A., & Dulmus, C. N. (2007). (Eds.). Handbook of Gerontology: Evidence-based Approaches to theory, Practice, & Policy. New York, NY: John Wiley.

Boron, J. B., Turiano, N. A., Willis, S. L., & Schaie, K. W. (2007a). Effects of cognitive training on change in accuracy in inductive reasoning ability. *Journals of Gerontology: Psychological Sciences, 62*(3), 179–186.

Boron, J. B., Willis, S. L., & Schaie, K. W. (2007b). Cognitive training gain as a predictor of mental status. *Journals of Gerontology: Psychological Sciences, 62*(1), 45–52.

Bosma, H., van Boxtel, M. P., Ponds, R. W., Jelicic, M., Houx, P., Metsemakers, J., & Jolles, J. (2003). "Engaged lifestyle and cognitive function in middle- and old-aged, non-demented persons: A reciprocal association?" *Zeitschrift für Gerontologie und Geriatrie, 35*(6), 575–581.

Boyke, J., Driemeyer, J., Gaser, C., Buechel, C., & May, A. (2008). Training-induced brain structure changes in the elderly. *Journal of Neuroscience, 28*(28), 7031–7035. doi:10.1523/JNEUROSCI.0742–08.2008.

Brehmer, Y., Li, S. C., Müller, V., von Oertzen, T., & Lindenberger, U. (2007). Memory plasticity across the lifespan: Developmental .*Psychology, 43*, 465–478.

Brehmer, Y., Li, S. C., Straube, B., Stoll, G., von Oertzen, T., Mueller, V., & Lindenberger, U. (2008). Comparing memory skill maintenance across the life span: Preservation in adults, increase in children. *Psychology & Aging, 23*(2), 227–238. doi:10.1037/0882–7974.23.2.227

Bretz, R. D., Jr., & Thompsett, R. E. (1992). Comparing traditional and integrative learning methods in organizational training programs. *Journal of Applied Psychology, 77*(6), 941–951.

Brickman, A. M., & Stern, Y. (2009). Aging and memory in humans. *Encyclopedia of Neuroscience, 1*, 175–180.

Brooks, J. O., III, Friedman, L., Pearman, A. M., Gray, C., & Yesavage, J. A. (1999a). Mnemonic training in older adults: Effects of age, length of training, and type of cognitive pretraining. *International Psychogeriatrics, 11*(1), 75–84.

Brooks, J. O., Friedman, L., & Yesavage, J. A. (1999b). Mnemonic training in older adults: Effects of age, length of training, and type of cognitive pretraining. *International Psychogeriatrics, 11*(1), 75–84.

Brown, C. H., & Liao, J. (1999). Principles for designing randomized preventive trials in mental health: An emerging developmental epidemiologic perspective. *American Journal of Community Psychology, 27*(5), 673–709.

Brown, J., Cooper-Kuhn, C. M., Kemperman, G., van Praag, H., Winkler, J., Gage, F. H., Kuhn, H. G. (2003). Enriched environment and physical activity stimulate hippocampal but not olfactory bulb neurogenesis. *European Journal of Neuroscience, 17*, 2042–2046.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical Linear Models*. Newbury Park, CA: Sage.

Buckner, R. L. (2004). Memory and executive function in ageing and AD: Multiple factors that cause decline and reserve factors that compensate. *Neuron*, *44*(1), 195–208.

Buitenweg, J. I. V., Murre, J. M. J., & Ridderinkhof, K. R. (2012). Brain training in progress: A review of trainability in healthy seniors. *Frontiers of Human Neuroscience 6*, 183. doi:10.3389/fnhum.2012.00183

Buiza, C., Etxeberria, I., Galdona, N., González, M. F., Arriola, E., de Munain, A. L., ... Yanguas, J. J. (2008). A randomized, two-year study of the efficacy of cognitive intervention on elderly people: The Donostia Longitudinal Study. *International Journal of Geriatric Psychiatry, 23*(1), 85–94.

Buschert, V., Bokde, A. L., & Hampel, H. (2010). Cognitive intervention in Alzheimer disease. *Nature Reviews Neurology*, *6*(9), 508–517. doi:10.1038/nrneurol.2010.113

Buschkuehl, M., Jaeggi, S. M., Hutchison, S., Perrig-Chiello, P., Däpp, C., Müller, M., Breil, F., Hoppeler, H., & Perrig, W. J. (2008). Impact of working memory training on memory performance in old-old adults. *Psychology & Aging 23,* 743. doi:10.1037/a0014342

Butler, R. (1991). Foreword. In J. Birren & D. Deutchman (Eds.). *Guiding Autobiography Groups for Older Adults* (pp. 219–234). Baltimore, MD: John Hopkins University Press.

Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: The HAROLD model. *Psychology & Aging, 17*(1), 85–100.

Cabeza, R., Anderson, N. D., Locantore, J. K., & McIntosh, A. R. (2002). Aging gracefully: Compensatory brain activity in high-performing older adults. *Neuroimage, 17*, 1394–1402.

Calero, M. D., & Navarro, E. (2007). Cognitive plasticity as a modulating variable on the effects of memory training in elderly persons. *Archives of Clinical Neuropsychology*, *22*(1), 63–72.

Carey, M. (2007). *A Literature Review: Cognitive Ageing and Cognitive Enhancement for Older Adults.* Centre for Research on Ageing. Perth; Curtin University of Technology.

Carter, S. D. (2002). Matching training methods and factors of cognitive ability: A means to improve training outcomes. *Human Resource Development Quarterly*, *13*(1), 71–87.

Cervilla, J. A., Prince, M., Lovestone, S., Mann, A., & Joels, S. (2002). Long-term predictors of cognitive outcome in a cohort of older people with hypertension. *British Journal of Psychiatry*, *177*(1), 66–71. doi:10.1192/bjp.177.1.66

Chen, H., & Cohen, P. (2006). Using individual growth model to analyze the change in quality of life from adolescence to adulthood. *Health & Quality of Life Outcomes*, *4*(10) doi:10.1186/1477-7525-4-10 1-7

Cheng, Y., Wu, W., Feng, W., Wang, J., Chen, Y., Shen, Y., ..., & Li, C. (2012). The effects of multi-domain versus single-domain cognitive training in non-demented older people: a randomized controlled trial. *BMC Medicine*, *10*(1), 30.

Christensen, H., Korten, A. E., Jorm, A. F., Henderson, A. S., Jacomb, P. A., Rodgers, B., & Mackinnon, A. J. (1997). Education and decline in cognitive performance: Compensatory but not protective. *International Journal of Geriatric Psychiatry* *12*(3), 323–330. doi:10.1002/(SICI)1099–1166(199703)12:3<323::AID-GPS492>3.0.CO;2–N

Christensen, H., Korten, A., Jorm, A. F., Henderson, A. S., Scott, R., & Mackinnon, A. J. (1996). Activity levels and cognitive functioning in an elderly community sample. *Age & Ageing*, *25*(1), 72–80. http://ageing.oxfordjournals.org/content/25/1/72.full.pdf

Christensen, H., Mackinnon, A. J., Korten, A. E., Jorm, A. F., Henderson, A. S., Jacomb, P., & Rodgers, B. (1999). An analysis of diversity in the cognitive performance of elderly community dwellers: Individual differences in change scores as a function of age. *Psychology & Aging*, *14*(3), 365–379. doi:10.1037/0882–7974.14. 3.365

Chu, B. C., Millis, S., Arango-Lasprilla, J. C., Hanks, R., Novack, T., & Hart, T. (2007). Measuring recovery in new learning and memory following traumatic brain injury: A mixed-effects modeling approach. *Journal of Clinical &*

*Experimental Neuropsychology, 29*(6), 617–625. doi:10.1080/13803390600878893

Clare, L., & Woods, B. (2003). Cognitive rehabilitation and cognitive training for early-stage Alzheimer's disease and vascular dementia. *Cochrane Database of Systematic Reviews* 2003; (4) doi: 10.1002/14651858.CD003260. Art. No.: CD003260.

Clare, L., Woods, B., Moniz-Cook, E., Orrell, M., & Spector, A. (2003). Cognitive rehabilitation and cognitive training interventions targeting memory functioning in early-stage Alzheimer's disease and vascular dementia (Cochrane Review). In *The Cochrane Library*, Issue 4. Chichester, UK: John Wiley & Sons Ltd.

CogState Ltd. (2008). *Why CogState?* New Haven, CT: Cogstate Ltd. www.cogstate.com/go/ academic/introduction/why-cogstate

CogState Ltd. (2015). *Cognitive Tests.* https://cogstate.com/competerized-tests/ cognitive-tasks. Accessed 3 February, 2015.

Cohen, M. (1988). The Revised Conners Parent Rating Scale: Factor structure replication with a diversified clinical sample. *Journal of Abnormal Child Psychology*, *16*, 187–196.

Colcombe, S. J., Kramer, A. F., Erickson, K. I., & Scalf, P. (2005). The implications of cortical recruitment and brain morphology for individual differences in cognitive performance in ageing humans. *Psychology & Aging, 20*(3), 363–375.

Collie, A., Darby, D. G., Falleti, M. G., Silbert, B., & Maruff, P. (2002). Determining the extent of cognitive change following coronary surgery: An analysis of statistical procedures. *Annals of Thoracic Surgery, 73,* 2005–2011.

Collie, A., Maruff, P., Darby, D. G., & McStephen, M. (2003). The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. *Journal of International Neuropsychological Society, 9*(3), 419–428. doi:10.1017/S1355617703930074

Collins, L., & Horn, J. L. (1991). *Best Methods for the Analysis of Change*. Washington, DC: APA Press.

Colsher, P. L., & Wallace, R. B. (1991). Longitudinal application of cognitive function measures in a defined population of community-dwelling elders. *Annals of Epidemiology*, *1*(3), 215–230.

Connell, A. M., & Frye, A. (2006). Growth mixture modeling in developmental psychology: Overview and demonstration of heterogeneity in development trajectories of adolescent antisocial behaviour. *Infant & Child Development, 15*(6): 609–621.

Craik, F. I., & Broadbent, D. E. (1983). On the transfer of information from temporary to permanent memory [and discussion]. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences*, *302*(1110), 341–359.

Craik, F.I. Byrd, M., & Swanson, J.M. (1987). Patterns of memory loss in three elderly samples. *Psychology of Aging 2*(1), 79–86.

Craik F. I. M., Winocur, G., Palmer, H., Binns, M. A., Edwards, M., Bridges, K., … Stuss, D. T. (2007). Cognitive rehabilitation in the elderly: Effects on memory. *Journal of International Neuropsychological Society*, *13*(1), 132–142.

Crawford, J. R., & Henry, J. D. (2003). The Depression Anxiety Stress Scales (DASS): Normative data and latent structure in a large non-clinical sample. *British Journal of Clinical Psychology, 42*(2), 111–131. doi:10.1348/014466503321903544

Cronbach, L. J., & Furby, L. (1970). How we should measure" change": Or should we? *Psychological Bulletin 74*(1), 68.

Crook, J. M., Kisvarday, Z. F., & Eysel, U. T. (1998). Evidence for a contribution of lateral inhibition to orientation tuning and direction selectivity in cat visual cortex: Reversible inactivation of functionally characterized sites combined with neuroanatomical tracing techniques. *European Journal of Neuroscience. 10*, 2056–2075.

Crowe, M., Andel, R., Pedersen, N.L., Johansson, B., & Gatz, M. (2003). Does participation in leisure activities lead to reduced risk of Alzheimer's disease? A prospective study of Swedish twins. *Journal of Gerontology: Psychological Sciences, 58B*, P249–P255.

Cuijpers, P., van Lier, P. A. C., van Straten, A., & Donker, M. (2005). Examining differential effects of psychological treatment of depressive disorder: An application of trajectory analyses. *Journal of Affective Disorders, 89*, 137–146.

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve Frequently Asked Questions About Growth Curve Modeling. *Journal of Cognition and Development : Official Journal of the Cognitive Development Society*, *11*(2), 121–136. http://doi.org/10.1080/15248371003699969

Daffner, K. R. (2010). Promoting successful cognitive ageing: A comprehensive review. *Journal of Alzheimer's Disease, 19*(4):1101–1122.

Dahlin, E., Stigsdotter-Neely, A. S., Larsson, A., Bäckman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science, 320*(5882), 1510–1512.

De Paula, J. J., Melo, L. P. C., Nicolato, R., Moraes, E. N., Bicalho, M. A., Hamdan, A. C., & Malloy-Diniz, L. F. (2012). "Reliability and construct validity of the

Rey –Auditory Verbal Learning Test in Brazilian elders”. *Revista de Psiquiatria Clínica* *39*(1), 19–23. www.scielo.br/scielo.php?script=sci_arttext&pid= S0101 -60832012000100004&lng=en&tlng=en.10.1590/S0101-60832012000100004

Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R. E, … Starr, J. M. (2009a). Age-associated cognitive decline. *British Medical Bulletin*, *92*(1), 135–152.

Deary, I. J., Johnson, W., & Houlihan, L. M. (2009b). Genetic foundations of human intelligence. *Human Genetics 126*, 215–232.

deFrias, C. M., Lövdén, M., Lindenberger, U., & Nilsson, L.-G. (2007). Revisiting the dedifferentiation hypothesis with longitudinal multicohort data. *Intelligence, 35*, 381–392.

Denney, N. W. (1984). A model of cognitive development across the life span. *Developmental Review, 4*(2), 171–191.

Denney, N. W., & Heidrich, S. M. (1990). Training effects on Raven’s Progressive matrices in young, middle-aged, and elderly adults. *Psychology & Aging*, *5*(1), 144–5.

Depp, C. A., Harmell, A., & Vahia, I. V. (2012). Successful cognitive ageing. *Current Topics in Behavioural Neuroscience, 10,* 35–50 doi:10.1007/7854_2011_158

Der, G., & Deary, I. J. (2006). Age and sex differences in reaction time in adulthood: Results from the United Kingdom Health and Lifestyle Survey: Correction. *Psychology & Aging, 21*(1), 62–73.

Donders, A. R., van der Heijden, G. J., Stijnen, T., & Moons, K. G. *M* (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology, 59*(10), 1087–1091. doi:10.1016/j.jclinepi.2006.01.014

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Changes in grey matter induced by training. *Nature, 427*(6972), 311–312.

Draganski, B., Gaser, C., Kempermann, G., Kuhn, H. G., Winkler, J., Büchel, C., & May, A. (2006). Temporal and spatial dynamics of brain structure changes during extensive learning. *Journal of Neuroscience, 26*(23), 6314–6317.

D'Unger, A. V., Land, K. C., McCall, P. L., & Nagin, D. S. (1998). How many latent classes of delinquent/criminal? Results from mixed Poisson regression analyses of the London, Philadelphia, and Racine cohort studies. *American Journal of Sociology, 103*, 1593–1630.

Duncan, S. C., Duncan, T. E., & Strycker, L. A. (2000). Risk and protective factors influencing adolescent problem behavior: A multivariate latent growth curve analysis. *Annals of Behavioral Medicine, 22*, 103–109.

Duncan, T. E., & Duncan, S. C. (2009). The ABC's of LGM: An introductory guide to latent variable growth curve modeling. *Social & Personality Psychology Compass*, *3*(6), 979–991.

Duncan, T. E., Duncan, S. C., & Stryker, L. A. (2006). *An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues, and Applications* (2nd ed.). Mahwah, NJ: Erlbaum.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues, & Applications*. Mahwah, NJ: Lawrence Erlbaum.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Okut, H., & Li, F. (2002). *Growth Mixture Modeling of Adolescent Alcohol Use Data*, Eugene, OR: Oregon Research Institute.

Dunlosky, J., & Hertzog, C. (1998a). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in Educational Theory & Practice* (pp. 249–275). Mahwah, NJ: Erlbaum.

Dunlosky, J., & Hertzog, C. (1998b). Ageing and deficits in associative memory: What is the role of strategy production? *Psychology & Aging, 13,* 597–607.

Eggleston, E. P., Laub, J. H., & Sampson, R. J. (2004). Methodological sensitivities to latent class analysis of long-term criminal trajectories. *Journal of Quantitative Criminology, 20*(1):1–25.

Ehninger, D., & Kempermann, G. (2003). Regional effects of wheel running and environmental enrichment on cell genesis and microglia proliferation in the adult murine neocortex. *Cerebral Cortex, 13*(8), 845–851.

Einstein, G. O., & McDaniel, M. A. (2004). *Memory Fitness: A Guide for Successful Aging*. New Haven, CT: Yale University Press.

Ekstrom, R., French, J., Harman, H., & Derman, D. (1976). *Kit of Factor-referenced Cognitive Tests* (Revised ed.). Princeton, NJ: Educational Testing Service; 1976.

Ellis, K. A., Bush, A. I., Darby, D., De Fazio, D., Foster, J., Hudson, P., ... Ames, D. (2009). The Australian Imageing, Biomarkers and Lifestyle (AIBL) study of ageing: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *International Psychogeriatrics, 21*(4), 672–687. doi:10.1017/S1041610209009405

Erber, J. T., & Szuchman, L. T. (1996). Memory performance in relation to age, verbal ability, and activity. *Experimental Aging Research, 22*, 59–72.

Espe-Pfeifer, P., & Wachsler-Felder, J. (2002). *Neuropsychological Interpretation of Objective Psychological Tests (2000).* New York, NY: Kluwer Academic/Plenum Publishers.

Fairchild, J. K., Friedman, L., Rosen, A. C., & Yesavage, J. A. (2013). Which older adults maintain benefit from cognitive training? Use of signal detection methods to identify long-term treatment gains. *International Psychogeriatrics 25*(4), 607–616. doi:10.1017/S1041610212002049

Falleti, M. G., Maruff, P., Collie P., & Darby, D. G. (2006). Practice effects associated with the repeated assessment of cognitive function using the CogState Battery at 10-minute, one-week and one-month test-retest intervals. *Journal of Clinical & Experimental Neuropsychology, 28*(7), 1095–1112. doi:10.1080/ 13803390500205718

Fandakova, Y., Shing, Y. L., & Lindenberger, U. (2012). Heterogeneity in memory training improvement among older adults: A latent class analysis. *Memory, 20*(6), 554–567. doi:10.1080/09658211.2012.687051

Feldman, D. E., Nicoll, R. A., & Malenka, R. C. (1999). Synaptic plasticity at thalamocortical synapses in developing rat somatosensory cortex: LTP, LTD, and silent synapses. *Journal of Neurobiology*, *41*(1), 92–101.

Feldman, B., Masyn, K., & Conger, R. (2009). New approaches to studying problem behaviors: A comparison of methods for modeling longitudinal, categorical data. *Developmental Psychology*, 45(3), 652–676. doi:10.1037/a0014851

Ferguson, G. A. (1956). On transfer and the abilities of man. *Canadian Journal of Psychology*, 10, 121–131.

Fillenbaum, G. G., George, L. K., & Palmore, E. B. (1985). Determinants and consequences of retirement among men of different races and economic levels. *Journals of Gerontology 40*, 85–94.

Fillit, H. M., Butler, R. N., O'Connell, A. W., Albert, M. S., Birren, J. E., Cotman, C. W., ... Tully, T. (2002, July). Achieving and maintaining cognitive vitality with ageing. *Mayo Clinic Proceedings, 77*(7), 681–696.

Floyd, M., & Scogin, F. (1997). Effects of memory training on the subjective memory functioning and mental health of older adults: A meta-analysis. *Psychology & Aging, 12*(1), 150–161.

Fogel, J.F., Hyman, R. B., Rock, B., Wolk-Klein, G. (2000). Predictors of hospital length of stay and nursing home placement in an elderly medical population. *Journal of the American Medical Directors Association*, *1*, 202–210.

Fratiglioni, L., Paillard-Borg, S., & Winblad, B. (2004). An active and socially integrated lifestyle in late life might protect against dementia. *Lancet Neurology, 3*(6), 343–353.

Fredrickson, J., Maruff, P., Woodward, M., Moore, L. Fredrickson, A., Sach, J., & Darby, D. (2010). Evaluation of the usability of a brief computerized cognitive screening test in older people for epidemiological studies. *Methods in Neuroepidemiology*, *34*(2), 65–75.

Frick, K. M., & Benoit, J. D. (2010). Use it or lose it: Environmental enrichment as a means to promote successful cognitive ageing. *Scientific World Journal, 10*, 1129–1141. doi:10.1100/tsw.2010.111

Garlick, D. (2002). Understanding the nature of the general factor of intelligence: the role of individual differences in neural plasticity as an explanatory mechanism. *Psychological Review*, *109*(1), 116.

Garrett, D. D., Macdonald, S. W., & Craik, F. I. (2012). Intraindividual reaction time variability is malleable: feedback-and education-related reductions in variability with age. *Frontiers in Human Neuroscience*, *6*, 101.

Gatz, M. (2005). Educating the brain to avoid dementia: Can mental exercise prevent Alzheimer disease? *PLoS Medicine 2*(1), e7.

Gehring, K., Aaronson, N. K., Gundy, C. M., Taphoorn, M. J., Sitskoorn, M. M. (2011). Predictors of neuropsychological improvement following cognitive rehabilitation in patients with gliomas. *Journal of the International Neuropsychology Society,* 17(2), 256–266. doi:10.1017/ S1355617710001530

Gehring, K., Sitskoorn, M. M., Gundy, C. M., Sikkes, S. A., Klein, M., Postma, T. J., … Aaronson, N. K. (2009). Cognitive rehabilitation in patients with gliomas: A randomized, controlled trial. *Journal of Clinical Oncology, 27*(22), 3712–3722. doi:10.1200/JCO.2008.20.5765

Gilsky, E. L., Rubin, S. R., & Davidson, P. S. R. (2001). Source memory in older adults: An encoding or retrieval problem? *Journal of Experimental Psychology: Learning, Memory & Cognition*, *27*(5), 1131–1146. doi:10.1037/0278–7393.27.5.1131

Gold, D. P., Andres, D., Etezadi, J., Arbuckle, T., Schwartzman, A., & Chaikelson, J. (1995). Structural equation model of intellectual change and continuity and predictors of intelligence in older men [published erratum appears in *Psychology & Aging* 1998, *13*(3), 434]. *Psychology & Aging*, *10*(2), 294–303.

Goldin, C. (1998). America’s graduation from high school: The evolution and spread of secondary schooling in the twentieth century. *Journal of Economic History, 58*(2), 345–374.

Gomez, R., Summers, M., Summers, A., Wolf, A., & Summers, J. J. (2014). Depression Anxiety Stress Scales-21: Factor structure and test-retest invariance, and temporal stability and uniqueness of latent factors. *Older Adults Journal of Psychopathology & Behavioral Assessment, 36*(2), 308–317.

Gontkovsky, S. T., Mold, J. W., & Beatty, W. W. (2002). Age and educational influences on RBANS index scores in a nondemented geriatric sample. *Clinical Neuropsychologist, 16*(3), 258–263.

Gould, E., & Gross, C. G. (2002). Neurogenesis in adult mammals: Some progress and problems. *The Journal of Neuroscience, 22*(3):619–623 doi:10.3389/fnins.2014.00045

Graf, P., & Uttl, B. (1995). Component processes of memory: Changes across the adult lifespan. *Swiss Journal of Psychology, 54*(2), 113–130.

Graham, J. E., Rockwood, K., Beattie, B. L., Eastwood, R., Gauthier, S., Tuokko, H., McDowell, I. (1997). Prevalence and severity of cognitive impairment with and without dementia in an elderly population. *Lancet, 349*, 1793–1796.

Graham, J. W., Taylor, B. J., & Cumsille, P. E. (2001). Planned missing data designs in analysis of change. In L. Collins, & A. Sayer (Eds.), *New Methods for the Analysis of Change* (pp. 335–353). Washington, DC: American Psychological Association.

Gravetter, F. J., & Wallnau, L. B. (2004). *Statistics for the Behavioral Sci*ences (6th ed.). Belmont, CA: Wadsworth.

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*, 534–537.

Green, R. E., Colella, B., Christensen, B., Johns, K., Frasca, D., Bayley, M., & Monette, G. (2008a). Examining moderators of cognitive recovery trajectories after moderate to severe traumatic brain injury. *Archives of Physical Medicine & Rehabilitation*, *89*(12), S16–S24. doi:10.1016/j.apmr.2008.09.551

Green, A. F., Rebok, G., & Lyketsos, C. G. (2008b). Influence of social network characteristics on cognition and functional status with ageing. *International Journal of Geriatric Psychiatry*, *23*(9), 972–978.

Greenbaum, P. E., Del Boca, F. K., Darkes, J., Wang, C., & Goldman, M. S. (2005). Variation in the drinking trajectories of freshman college students. *Journal of Consulting & Clinical Psychology*, *73*(2), 229–238.

Greene, L. S., & Williams, H. G (1996). Aging and coordination from the dynamic pattern perspective. In A. M. Ferrandez, & N. Teasdale (Eds.), *Change in Sensory Motor Behavior in Aging* (pp. 89–131). Holland: Elsevier Science BV.

Greenwood, P. M. (2007). Functional plasticity in cognitive ageing: Review and hypothesis. *Neuropsychology*, *21*(6), 657–653.

Greenwood, P. M., & Parasuraman, R. (2010). Neuronal and cognitive plasticity: A neurocognitive framework for ameliorating cognitive ageing. *Frontiers in Aging Neuroscience*, *2*, 150. doi:10.3389/fnagi.2010.00150

Griffiths, J., McCutcheon, C., Silbert, B., & Maruff, P. (2006). A prospective observational study of the effect of night duty on the cognitive function of anaesthetic registrars. *Anaesthesia & Intensive Care*, *34*(5), 621–628

Gross, A. L., & Rebok, G. W. (2011). Memory training and strategy use in older adults: Results from the ACTIVE Study. *Psychology & Aging*. *26*(3), 503–517. doi:10.1037/a0022687

Gross A.L., Rebok, G.W., Brandt, J., Tommet, D., Marsiske, M., & Jones, R.N., (2012). Modeling learning and memory using verbal learning tests: Results from ACTIVE. *Journals of Gerontology Series B: Psychological Sciences & Social Sciences*, 68(2), 153–167. doi:10.1093/geronb/gbs053

Gross, A. L., Rebok, G. W., Unverzagt, F. W., Willis, S. L., & Brandt, J. (2011). Word list memory predicts everyday function and problem-solving in the elderly: Results from the ACTIVE cognitive intervention trial. *Aging, Neuropsychology, & Cognition*, *18*(2), 129–146.

Gueorguieva, R., Wu, R., Pittman. B., Cramer. J., Rosenheck, R. A., O'malley, S. S., & Krystal JH. (2007). New insights into the efficacy of naltrexone based on trajectory-based reanalyses of two negative clinical trials. *Biological Psychiatry 61*(11), 1290–1295.

Gur, R. C., Mozley, P. D., Resnick, S. M., Gottlieb, G. L., Kohn, M., Zimmerman, R., ... Berretta, D. (1991). Gender differences in age effect on brain atrophy measured by magnetic resonance imageing. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(7), 2845–2849.

Habib, R., Nyberg, L., Nilsson, L. G. (2007). Cognitive and non-cognitive factors contributing to the longitudinal identification of successful older adults in the Betula study. *Aging Neuropsychology & Cognition. 14*, 257–273.

Hallett, M. (2001). Brain plasticity and recovery from hemiplegia. *Journal of Medical Speech-Language Pathology, 9*(2), 107–115.

Hastings, E. C., & West, R. L. (2009). The relative success of a group-based and self-help memory training program, *Psychology & Aging*, *24*(3), 586–594. doi:10.1037/ a0016951

Hedden, T., & Gabrieli, J. D. E. (2004). Insights into the ageing mind: A view from cognitive neuroscience. *Nature, 5*(2), 87–96.

Hedden, T., & Gabrieli, J. D. E (2005). Healthy and pathological processes in adult development: New evidence from neuroimaging of the ageing brain. *Current Opinion in Neurology, 18*(6), 740–747.

Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance following focal cortical lesions. *Neuropsychology, 18*, 284–295.

Herlitz, A., Nilsson, L. G., & Bäckman, L. (1997). Gender differences in episodic memory. *Memory & Cognition*, *25*(6), 801–811.

Hertzog, C., & Dunlosky, J. (2004). Aging, metacognition, and cognitive control. *The Psychology of Learning & Motivation, 45,* 215–251.

Hertzog, C., Kramer, A. F., Wilson, R. S., & Lindenberger, U. (2008a). Enrichment effects on adult cognitive development can the functional capacity of older adults be preserved and enhanced? *Psychological Science in the Public Interest*, *9*(1), 1–65.

Hertzog, C., Kramer, A. F., Wilson, R. S., & Lindenberger, U. (2009). Enrichment effects on adult cognitive development: Can the functional capacity of older adults be preserved and enhanced? *Journal of the Association for Psychological Science, 9*(1), 1–65.

Hertzog, C., Price, J., & Dunlosky, J. (2008b). How is knowledge generated about memory encoding strategy effectiveness? *Learning & Individual Differences, 18*(4), 430–445.

Hertzog, C., & Robinson, A. E. (2005). Metacognition and intelligence. In O. Wilhelm, & R. W. Engle (Eds.), *Understanding & Measuring Intelligence* (pp. 101–123). London, UK: Sage.

Hess, T. (2005). Memory and ageing in context. *Psychological Bulletin*, *131*(3), 383–406.f

Hill, R. D., & Bäckman, L. (2000). Theoretical and methodological issues in memory training. In R. D. Hill, L. Bäckman, & A. Stigsdotter-Neely (Eds.), *Cognitive Rehabilitation in Old Age* (pp. 23–41). New York, NY: Oxford University Press.

Hill, R. D., Bäckman, L., &. Stigsdotter-Neely, A. (Eds.) (2000). *Cognitive Rehabilitation in Old Age*. New York, NY: Oxford University Press.

Hill, R. D., Wahlin, Å., Winblad, B., & Bäckman, L. (1995). The role of demographic and lifestyle variables in utilizing cognitive support for episodic remembering among very old adults. *Journals of Gerontology: Psychological Sciences*, *50*(4), 219–227.

Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods, 11*, 36–53.

Hirshson, C (2010). Gender differences in neuropsychological test performance in non-demented older adults. *Journal of International Neuropsychological Society 16* (Suppl 1), 219–234.

Hix-Small, H., Duncan, T. E., Duncan, S. C., & Okut, H. (2004). A multivariate associative finite growth mixture modeling approach examining adolescent alcohol and marijuana use. *Journal of Psychopathology & Behavioral Associations*, *26*(4), 255–270.

Hohaus, L. (2007). Remembering to age successfully: Evaluation of a successful ageing approach to memory enhancement. *International Psychogeriatrics*, 19(1), 137–150.

Holdnack, J. A. (2001). *WTAR: Wechsler Test of Adult Reading Manual*. San Antonio, TX: Psychological Corporation.

Hooyman, N. A., & Kiyak, H. A. (2002). *Social Gerontology: A Multidisciplinary Perspective*. Boston, MA: Allyn & Bacon.

Huang, D., Brecht, M. L., Hara, M., & Hser, Y. I. (2010). Influences of covariates on growth mixture modeling. *Journal of Drug Issues*, *40*(1), 173–194.

Hultsch, D. F., Hammer, M., & Small, B. J. (1993). Age differences in cognitive performance in later life: Relationships to selfreported health and activity life style. *Journals of Gerontology, 48*, 1–11.

Hultsch, D. F., Hertzog, C., & Dixon, R. A. (1990). Ability correlates of memory performance in adulthood and ageing. *Psychology & Aging, 5,* 356–368.

Hultsch, D. F., Hertzog, C., Dixon, R. A., & Small, B. J. (1998). *Memory Change in the Aged.* New York, NY: Cambridge University Press.

Hultsch, D. F., Hertzog, C., Small, B. J., & Dixon, R. A. (1999). Use it or lose it: engaged lifestyle as a buffer of cognitive decline in ageing? *Psychology & Aging, 14*(2), 245–263.

Hunt, E. (1978). Mechanics of verbal ability. *Psychological Review, 85*(2), 109–130.

Hyde, M., Wiggins, R., Higgs, P., & Blane, D. (2003). A measure of quality of life in early old age: The theory, development and properties of a needs satisfaction model (CASP-19). *Aging & Mental Health,* **7(3),** 186–194. doi:10.1080/ 1360786031000101157

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(19), 6829–6833.

Jarvik, L. F., & Falek, A. (1963). Intellectual stability and survival in the aged. *Journal of Gerontology*, *18*, 173–176.

Jennings, J. M., & Jacoby, L. L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology & Aging, 8,* 283–293.

Jessberger, S., & Gage, F. H. (2008). Stem-cell-associated structural and functional plasticity in the ageing hippocampus. *Psychology & Aging*, *23*(4), 684.

Johnson, M., Bengsteon, V., Coleman, P., & Kirkwood, B. (2005). *The Cambridge Handbook of Age and Ageing.* Cambridge: Cambridge University Press.

Johnson-Greene, D. (2004). Test review: Dementia Rating Scale-2 (DRS-2). *Archives of Clinical Neuropsychology*, *19*(1), 145–147.

Johnstone, B., Vieth, A. Z., Johnson, J. C., & Shaw, J. A. (2000). Recall as a function of single versus multiple trials: Implications for rehabilitation. *Rehabilitation Psychology*, *45*(1), 3.

Jolles, D. D., & Crone, E. A. (2012) Training the developing brain: a neurocognitive perspective. *Frontiers of Human Neuroscience, 6*:76. doi:10.3389/fnhum.2012.00076

Jones, R.N., Manly, J. Glymour, M. M., Rentz, D.M., Jefferson, A. L., & Stern, Y. (2011). Conceptual and measurement challenges in research on cognitive reserve. *Journal of the International Neuropsychological Society,* 17, 593–601. doi:10.1017/S1355617710001748

Jones, S., Nyberg, L., Sandblom, J., Stigsdotter-Neely, A., Ingvar, M., Petersson, K., & Bäckman, L. (2006). Cognitive and neural plasticity in ageing: General and task-specific limitations. *Neuroscience & Biobehavioral Reviews*, *30*(6), 864–871. doi:10.1016/j.neubiorev.2006.06.012

Jones, R. N., Rosenberg, A. L., Morris, J. N., Allaire, J. C., McCoy, K. J. M., Marsiske, M., … Malloy, P. F. (2005). A growth curve model of learning acquisition among cognitively normal older adults. *Experimental Aging Research, 31*(3), 291–312. doi:10.1080/03610730590948195

Jopp, D., & Hertzog, C. (2007). Activities, self-referent memory beliefs, and cognitive performance: Evidence for direct and mediated effects. *Psychology & Aging, 22*(4), 811–825.

Jöreskog, K.G., & Sörbom, D. (1979). *Advances in Factor Analysis and Structural Equation Models.* Cambridge, MA: Abt Books.

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social & Personality Psychology Compass, 2*, 302–317. doi:10.1111/j.1751-9004.2007.00054.x

Kaplan, D. (Ed.). (2004). *The Sage Handbook of Quantitative Methodology for the Social Sciences*. Thousand Oaks, CA: Sage Publications.

Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task – switching training. *Developmental Science, 12*, 978–990.

Karbach**,** J., & Schubert, T. (2013). Training-induced cognitive and neural plasticity**.** *Frontiers in Human Neuroscience, 7*(48)**,** 1–22. doi:10.3389/fnhum.2013.00048

Karlsson, T., Backman, L., Herlitz, A., Nilsson, L.-G., Winblad, B., & Osterlind, P.-O. (1989). Memory improvement at different stages of Alzheimer's disease. *Neuropsychologia, 27*(5), 737–742.

Karp, A., Andel, R., Parker, M. G., Wang, H-X., Winblad, B., & Fratiglioni, L. (2009). Mentally stimulating activities at work during midlife and dementia risk after age 75: Follow-up study from the Kungsholmen Project. *American Journal of Geriatric Psychiatry, 17*(3). 227–236. doi:10.1097/JGP.0b013e318190b691

Karp, A., Paillard-Borg, S., Wang, H., Silverstein, M., Winblad, B., & Fratiglioni, L. (2006). Mental, physical and social components in leisure activities equally contribute to decreased dementia risk. *Dementia & Geriatric Cognitive Disorders, 21*(2), 65–73.

Katz, L. C., & Rubin, M. (1999). *Keeping Your Brain Alive*. New York, NY: Workman Publishing.

Katzman, R. (1993). Education and the prevalence of dementia and Alzheimer's disease. *Neurology, 43*(13), 13–20.

Keller, M. B. (2001). Long-term treatment of recurrent and chronic depression. *Psychiatric Services, 62*(Suppl. 24), 3–5.

Kempermann, G., Gast, D., & Gage, F. H. (2002). Neuroplasticity in old age: Sustained five-fold induction of hippocampal neurogenesis by long-term environmental enrichment. *Annals of Neurology, 52*, 135–143.

Kempermann, G., Kuhn, H. G., & Gage, F. H. (1998). Experience-induced neurogenesis in the senescent dentate gyrus. *Journal of Neuroscience, 18*(9), 3206–3212.

Klekociuk, S. Z., & Summers, M. J. (2013). Self-fulfilling prophecy of episodic memory impairment in mild cognitive impairment: Do episodic memory deficits identified at classification remain evident when later examined with different memory tests? *Neurology Research International,* 1–6 doi:10.1155/2013/437013

Kliegl, R., Smith, J., & Baltes, P. B. (1989). Testing-the-limits and the study of adult age differences in cognitive plasticity of a mnemonic skill. *Developmental Psychology, 25*, 247–256.

Kliegl, R., Smith, J., & Baltes, P. B. (1990). On the locus and process of magnification of age differences during mnemonic training. *Developmental Psychology 26*(6), 894–904.

Kral, V. A. (1962). Senescent forgetfulness: Benign and malignant. *Canadian Medical Association Journal, 86*(6), 257–260.

Kramer, A. F., Bherer, L., Colcombe, S. J., Dong, W., & Greenough, W. T. (2004). Environmental influences on cognitive and brain plasticity during ageing. *Journals of Gerontology: Biological Sciences & Medical Sciences, 59*(9), M940–M957.

Kramer, M. A., Eden, U. T., Cash, S. S., & Kolaczyk, E. D. (2009). Network inference with confidence from multivariate time series. *Physical Review E, 79*, 061916.

Kramer, A. F., Larish, J. F., & Strayer, D. L. (1995). Training for attentional control in dual-task settings – a comparison of young and old adults. *Journal of Experimental Psychology Applications, 1*(1), 50–76.

Kramer, A. F., & Willis, S. (2002). Enhancing the cognitive vitality of older adults. *Current Directions in Psychological Science, 11*(5), 173–177.

Kramer, A. F., & Willis, S. (2003). Cognitive plasticity and ageing. In B. Ross (Ed.), *Psychology of Learning & Motivation* (*43*, pp. 267–302). New York, NY: Academic Press.

Krech, D., Rosenzweig, M. R., & Bennett, E. L. (1956). Effects of environmental complexity and training on brain chemistry. *Journal of Comparative & Physiological Psychology, 53*(6), 509–519.

Kreuter, F., & Muthén, B. (2007). Longitudinal modeling of population heterogeneity: Methodological challenges to the analysis of empirically derived criminal trajectory profiles. In G. R. Hancock, & K. M. Samuelsen (Eds.), *Advances in Latent Variable Mixture Models* (pp. 53–75). Charlotte, NC: Information Age Publishing.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38*(4), 963–974.

Langbaum, J. B. S., Rebok, G. W., Bandeen-Roche, K., & Carlson, M. C. (2009). Predicting memory training response patterns: Results from ACTIVE. *Journals of Gerontology: Psychological Sciences, 64B(1)*, 14–23. doi:10.1093/geronb/ gbn026

Larrabee, G. J., & Crook, T. H. (1993). Do men show more rapid age-associated decline in simulated everyday verbal memory than do women? *Psychology & Aging, 8*(1), 68–71.

LaRue, A. (2010). Healthy brain ageing: Role of cognitive reserve, cognitive stimulation, and cognitive exercises. *Clinics in Geriatric Medicine*, *26*(1), 99–111. doi:10.1016/j.cger.2009.11.003

Lautrey, J. (2003). A pluralistic approach to cognitive differentiation and development. In R. J. Sternberg, J. Lautrey, & T. I. Lubart (Eds.), *Models of Intelligence: International Perspectives* (pp. 117–131). Washington, DC: American Psychological Association.

Lawton, M. P. (1982). Competence, environmental press and the adaptation of older people. In M. P. Lawton, P. G. Windley, & T. O. Bverts. Eds., *Aging and the Environment: Theoretical Approaches*, pp.33–59. New York: Springer.

Leoutsakos, J. M. S., Muthén, B. O., Breitner, J., & Lyketsos, C. G. (2012). Effects of non-steroidal anti-inflammatory drug treatments on cognitive decline vary by phase of pre-clinical Alzheimer's disease: Findings from the randomized controlled Alzheimer's Disease Anti-inflammatory Prevention Trial. *International Journal of Geriatric Psychiatry*, *27*(4), 364–374. doi:10.1002/gps.2723

Levy, R. (1994). Aging-associated cognitive decline. *International Psychogeriatrics*, *6*(1), 63–68.

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological Assessment*. New York, NY: Oxford University Press.

Li, F., & Acock, A. C. (1999). *Latent Curve Analysis: A Manual for Research Data Analysis.* Presented at the National Council on Family Relations, Irvine, CA. http://oregonstate.edu/dept/hdfs/ papers/lgcmanual.pdf

Li, F., Duncan, T. E., Duncan, S. C., & Acock, A. (2001). Latent growth modeling of longitudinal data: A finite growth mixture modeling approach. *Structural Equation Modeling*, *8*(4), 493–530. doi:10.1207/S15328007SEM0804_01

Li, S.-C. (2003). Biocultural orchestration of developmental plasticity across levels: The interplay of biology and culture in shaping the mind and behavior across the lifespan. *Psychological Bulletin*, *129*, 171–194.

Lieberman, M. A. (1965). Psychological correlates of impending death: Some preliminary observations. *Journal of Gerontology*, *20*, 181–190.

Lindenberger, U., & Baltes, P. B. (1997). Intellectual functioning in old and very old age: Cross-sectional results from the Berlin Aging Study. *Psychology & Aging*, *12*, 410–432.

Lindenberger, U., & Ghisletta, P. (2009). Cognitive and sensory declines in old age: Gauging the evidence for a common cause. *Psychology & Ageing*, *24*, 1–16. doi:10.1037/a0014986

Lindenberger, U., Nagel, I. E., Chicherio, C., Li, S. C., Heekeren, H. R., & Bäckman, L. (2008). Age-related decline in brain resources modulates genetic effects on cognitive functioning. *Frontiers in Neuroscience, 2*(2), 234–244. doi:10.3389/neuro.01.039.2008

Lindenberger, U., & von Oertzen, T. (2006). Variability in cognitive ageing: From taxonomy to theory. In E. Bialystok, & F. I. M. Craik (Eds.), *Lifespan Cognition: Mechanisms of Change* (pp. 297–314). Oxford, UK: Oxford University Press.

Little, R. J. (1995). Modeling the dropout mechanism in repeated-measures studies. *Journal of American Statistical Association, 90*(431), 1112–1121.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical Analysis With Missing Data.* New York, NY: Wiley & Sons.

Little, R.J., & Rubin, D.B. (2002). *Statistical Analysis With Missing Data* (2nd ed.). New York: John Wiley & Sons.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrik*a, 88, 767–778. doi:10.1093/biomet/88.3.767

Longabaugh, R., Wirtz, P. W., Zweben, A., & Stout, R. (2001). Network support for drinking: Project MATCH hypotheses: Results and causal chain analyses. In R. Longabaugh, & P. W. Wirtz, (Eds.), *NIAAA Project MATCH Monograph Series* (*8*, pp. 260–275), NIH Publication no. 01–4238. Washington, DC: Government Printing Office.

Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., & Schmiedek, F. (2010). A theoretical framework for the study of adult cognitive plasticity, *Psychological Bulletin 136*(4), 659–676. doi:10.137/a0020080

Lövdén, M., Brehmer, Y., Li, S.-C., & Lindenberger, U. (2012). Training- induced compensation versus magnification of individual differences in memory performance. *Frontiers of Human Neurosciences 6*, 141. doi:10.3389/ fnhum.2012.00141

Lövdén, M., Rőnnlund, M., Wahlin, A., Backman, L., Nyberg, L., & Nilsson, L. G. (2004). The extent of stability and change in episodic and semantic memory in old age: Demographic predictors of level and change. *Journals of Gerontology: Series B Psychological Sciences*, *59*(3), P130–P134.

Lovelace, E. A., & Twohig, P. T. (1990). Healthy older adults' perceptions of their memory functioning and use of mnemonics. *Bulletin of the Psychonomic Society*, *28*(2), 115–118.

Lovibond, S. H., & Lovibond, P. F. (1995). *Depression Anxiety Stress Scale* (2nd ed.). Sydney, NSW: Psychology Foundation.

Lubke, G., & Muthén, B. O. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling*, *14*(1), 26–47.

Lupton, M. K., Stahl, D., Archer, N., Foy, C., Poppe, M., Lovestone, S., ... Powell, J. F. (2010). Education, occupation and retirement age effects on the age of onset of Alzheimer's disease. *International Journal of Geriatric Psychiatry*, *25*(1), 30–36. doi:10.1002/gps.2294

Lustig, C., & Flegal, K. E. (2008). Targeting latent function: Encouraging effective encoding for successful memory training and transfer. *Psychology & Aging, 23*(4):754–764.

Lustig, C., Shah, P., Seidler, R., & Reuter-Lorenz, P. A. (2009). Aging, training, and the brain: A review and future directions. *Neuropsychological Review, 19*(4), 504–522. doi:10.1007/s11065–009–9119–9

Luszcz, M., Bryan, J., & Kent, P. (1997). Predicting episodic memory performance of very old men and women: contributions from age, depression, activity, cognitive ability, and speed. *Psychology & Aging*, *12*(2), 340.

Lutz, W., Saunders, S. M., Leon, S. C., Martinovich, Z., Kosfelder, J., Schulte, D., ... Tholen, S. (2006). Empirically and clinically useful decision making in psychotherapy: Differential predictions with treatment response models. *Psychological Assessment*, 18, 133–141.

MacCallum, R. C., Kim, C., Malarkey, W. B., & Kiecolt-Glaser, J. K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research, 32*, 215–253.

Macdonald, S. W., Hultsch, D. F., & Dixon, R. A. (2003). Performance variability is related to change in cognition: evidence from the Victoria Longitudinal Study. *Psychology & Aging*, *18*(3), 510.

Magalhães, S. S., & Hamdan, A. C. (2010). Influence of verbal intelligence, sex, age and education on the Rey Auditory Verbal Learning Test. *Psychology & Neuroscience, 3*(1), 85–91. doi:10.3922/j.psns.2010.1.011

Magalhães, S. S., Malloy-Diniz, L. F., & Hamdan, A. C. (2012). Validity convergent and reliability test-retest of the Rey Auditory Verbal Learning Test. *Clinical Neuropsychiatry, 9*(3), 129–137.

Mahncke, H. W., Bronstone, A., & Merzenchi, M. M. (2006a). Brain plasticity and functional losses in the aged: A scientific base for a novel intervention. In A.

R. Moller (Ed.), *Progress in Brain Research: Reprogramming the Brain* (*157*, pp. 81–109). Texas: Elsevier.

Mahncke, H. W., Connor, B. B., Appelman, J., Ahsanuddin, O. N., Hardy, J. L., Wood, R. A., ... Merzenich, M. M. (2006b). Memory enhancement in healthy older adults using a brain plasticity-based training program: A randomized, controlled study. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(33), 12523–12528.

Makdissi, M., Collie, A., Maruff, P., Darby, D.G., Bush, A., McCrory, P., & Bennell, K. (2001). Computerised cognitive assessment of concussed Australian Rules footballers. *British Journal of Sports Medicine, 35,* 354–360.

Martin, M., Clare, L., Altgassen, A. M., Cameron, M. H., Zehnder, F. (2011). Cognition-based interventions for healthy older people and people with mild cognitive impairment. *Cochrane Database of Systematic Reviews.* 2011; (1) doi:10.1002/14651858. CD006220.pub2. Art No.: CD006220

Maruff, P., Thomas, E., Cysique, L., Brew, B., Collie, A., Snyder, P., & Pietrzak, R. H. (2009). Validity of the CogState brief battery: Relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, schizophrenia, and AIDS dementia complex. *Archives of Clinical Neuropsychology*, *24*(2), 165–178.

Mattis, S., Jurica, P. J., & Leitten, C. L. (2001). *Dementia Rating Scale-2*. Lutz, FL: Psychological Assessment Resources.

Mattson, M. P., Duan, W., Chan, S. L., Cheng, A., Haughey, N., Gary, D. S., ... Furukawa, K. (2002). Neuroprotective and neurorestorative signal transduction mechanisms in brain ageing: Modification by genes, diet and behavior. *Neurobiology of Aging*, *23*(5), 695–705.

McCutcheon, A., & Mills, C. (2007). Categorical data analysis: Log-linear and latent class models. In E. Scarborough, & E. Tanenbaum (Eds.), *Research Strategies in the Social Sciences: A Guide to New Approaches* (pp. 71–94). Oxford, UK: Oxford University Press.

McDougall, G. J. (1999). Cognitive interventions among older adults. *Annual Review of Nursing Research, 17*(1), 757–778.

McKelvie, S. J., Standing, L., St Jean, D., & Law, J. (1993). Gender differences in recognition memory for faces and cars: Evidence for the interest hypothesis. *Bulletin of the Psychonomic Society, 31*, 447–448.

McKitrick, L. A., Friedman, L. F., Brooks, J. O., Pearman, A., Kraemer, H. C., & Yesavage, J. A. (1999). Predicting response of older adults to mnemonic training: who will benefit? *International Psychogeriatrics, 11*(3), 289–300. doi:10.1017/S1041610299005852

McLachlan, G., & Peel, D. (2000). *Finite Mixture Models.* New York: Wiley.

McLellan, D. L. (1991). Functional recovery and the principles of disability medicine. In M. Swash & J. Oxbury (Eds.), *Clinical Neurology* (pp. 768–790). Edinburgh: Churchill Livingstone.

Medalia, A., & Richardson, R. (2005). What predicts a good response to cognitive remediation interventions? *Schizophrenia Bulletin, 31*(4), 942–953.

Meinz, E. J., & Salthouse, T. A. (1998). Is age kinder to females than males? *Psychonomic Bulletin & Review, 5*(1), 56–64.

Meredith, W., & Tisak, J. (1984). "*Tuckerizing" curves*. Paper presented at the Annual meeting of the Psychometric Society, Santa Barbara, CA.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*(1), 107–122.

Mitchell, M. B., Cimino, C. R., Benitez, A., Brown, C. L., Gibbons, L. E., Kennison, R. F., ... Piccinin, A. M. (2012). Cognitively stimulating activities: Effects on cognition across four studies with up to 21 years of longitudinal data. *Journal of Aging Research*, vol. 2012, Article ID 461592, 2012. doi:10.1155/2012/461592

Mohs, R. C., Ashman, T. A., Jantzen, K., Albert, M., Brandt, J., Gordon, B., ... Stern, Y. (1998). A study of the efficacy of a comprehensive memory enhancement program in healthy elderly persons. *Psychiatry Research*, *77*(3), 183–195.

Morse, C. K. (1993). Does variability increase with age? An archival study of cognitive measures. *Psychology & Aging*, *8*(2), 156.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*(1), 105–125.

Mortimer, J. A. (1997). Brain reserve and the clinical expression of Alzheimer's disease. *Geriatrics*, *52*, S50–S53.

Muthén, B. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R E. Schumacker (Eds.), *New Developments and Techniques in Structural Equation Modeling* (pp. 1–33). Mahwah, NJ: Erlbaum.

Muthén, B. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika, 29*, 81–117.

Muthen, B. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003), *Psychological Methods,* 8, 369-377.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of Quantitative Methodology for the Social Sciences* (pp. 345–368). Newbury Park, CA: Sage.

Muthén, B. (2005). *MPlus Discussion.* www.statmodel.com/discussion/messages/ 14/300.html? 1341951329

Muthén, B., Brown, C. H., Khoo, S., Yang, C., & Jo, B. (1998). *General Growth Mixture Modeling of Latent Trajectory Classes: Perspectives and Prospects*. Paper presented at the meeting of the Prevention Science and Methodology Group, Tempe, AZ.

Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S. T., Yang, C. … & Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics, 3*(4), 459–475.

Muthén, B., Jo, B, & Brown, C. H. (2003). Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City [with Comment]. *Journal of the American Statistical Association, 98*(462) 311–314.

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical & Experimental Research*, *24*(6), 882–891.

Muthén, L. K., & Muthén, B. O. (1998–2010). *MPlus User's Guide* (6th ed.). Los Angeles, CA: Muthén and Muthén.

Muthén, L., & Muthén, B. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 4,* 599–620.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55*(2), 463–446.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods, 4*(2), 139–57.

Nagin, D. (2005). *Group-based Modeling of Development*. Cambridge, MA: Harvard University Press.

Nagin, D. S., & Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model. *Criminology, 31*(3), 327–362.

Nagin, D. S., & Odgers, C. L. (2010). Group based trajectory modeling in clinical research. *Annual Review of Clinical Psychology, 6,* 109–138. doi:10.1146/annurev.clinpsy.121208.131413

Nagin, D., Pagani, L., Tremblay, R. E., & Vitaro, F. (2003). Life course turning points: A case study of the effect of school failure on interpersonal violence. *Development & Psychopathology. 15*(2), 343–361. doi:10.1017/ S0954579403000191

Neils-Strunjas, J., Krikorian, R., Shidler, M., & Likoy, S. (2001). The influence of learning style and cognitive ability on recall of names and faces in an older population. *Journal of General Psychology, 128*(4), 433–445.

Nelson, E. A., & Dannefer, D. (1992). Aged heterogeneity: Fact or fiction? The fate of diversity in gerontological research. *Gerontologist*, *32*(1), 17–23.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The Psychology of Learning & Motivation* (*26*, pp. 125–141). New York, NY: Academic Press.

Neuropathology Group of the Medical Research Council Cognitive Function and Ageing Study (MRC CFAS). (2001). Pathological correlates of late-onset dementia in a multicentre, community-based population in England and Wales. MRC CFAS. *Lancet, 375*(9251), 169–175.

Nithianantharajah, J., & Hannan, A. J. (2006). Enriched environments, experience-dependent plasticity and disorders of the nervous system. *Nature Reviews Neuroscience, 7*(9), 697–709. doi:10.1038/nrn1970

Noack, H., Lövdén, M., Schmiedek, F., & Lindenberger, U. (2009). Cognitive plasticity in adulthood and old age: Gauging the generality of cognitive intervention effects. *Restorative Neurology & Neuroscience, 27*(5), 535–553.

Nyberg, L., Sandblom, J., Jones, S., Stigsdotter-Neely, A. S., Petersson, K. M., Ingvar, M., & Bäckman, L. (2003). Neural correlates of training-related memory improvement in adulthood and ageing. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(23), 13728–13733.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*, 535–569.

O'Hara, R., Brooks, J. O., Friedman, L., Schröder, C. M., Morgan, K. S., & Kraemer, H. C. (2007). Long-term effects of mnemonic training in community-dwelling older adults. *Journal of Psychiatric Research*, *41*(7), 585–590.

Olesen, P. J., Westerberg, H., & Klingberg, T., 2004. Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience,* 7(1), 75–79.

Oswald, W., Gunzelmann, T., Rupprecht, R., & Hagen, B. (2006). Differential effects of single versus combined cognitive and physical training with older adults: The SimA study in a five-year perspective. *European Journal of Ageing, 3*(4), 179–192.

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., ..., & Ballard, C. G. (2010). Putting brain training to the test. *Nature*, *465*(7299), 775–778.

Pallant, J. (2007). *SPSS Survival Analysis*. Crowes Nest, NSW: Allen & Unwin.

Papp, K. V., Walsh, S. J., & Snyder, P. J. (2009). Immediate and delayed effects of cognitive interventions in healthy elderly: A review of current literature and future directions. *Alzheimer's & Dementia, 5*(1), 50–60.

Park, D. C., Gutchess, A. H., Meade, M. L., & Stine-Morrow, E. A. (2007). Improving cognitive function in older adults: nontraditional approaches. *The Journals of Gerontology Series B: Psychological Sciences & Social Sciences*, *62*(Special Issue 1), 45–52.

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology & Aging*, *17*(2), 299–320.

Park, D. C., & Reuter-Lorenz, P. (2009). The adaptive brain: Aging and neurocognitive scaffolding. *Annual Review of Psychology* 60, 173–196. doi:10.1146/annurev. psych.59.103006.093656

Pervin, L. A. (1978). Theoretical approaches to the analysis of individual-environment interaction. In L. A. Pervin, & M. E. Lewis (Eds.), *Perspectives in Interactional Psychology* (pp. 67–86). New York, NY: Plenum Press.

Petras, H., & Masyn, K. (2010). General growth mixture analysis with antecedents and consequences of change. In A. R. Piquero, & D. Weisburd (Eds.), *Handbook of Quantitative Criminology* (pp. 69–100). New York, NY: Springer.

Pietrzak, R. H., Cohen, H., & Snyder, P. J. (2007). Spatial learning efficiency and error monitoring in normal ageing: An investigation using a novel hidden maze learning test. *Archives of Clinical Neuropsychology, 22*, 235–245.

Pietrzak, R. H., Lim, Y. Y., Neumeister, A., Ames, D., Ellis, K. A., Harrington, K., ..., & Maruff, P. (2015). Amyloid-β, anxiety, and cognitive decline in preclinical Alzheimer Disease: A multicenter, prospective cohort study. *Journal of the American Medical Association, Psychiatry, 72*, 284–291.

Pietrzak, R. H., Maruff, P., & Snyder, P. J. (2009). Methodological improvements in quantifying cognitive change in clinical trials: an example with single-dose administration of donepezil. *The Journal of Nutrition, Health & Aging, 13*(3), 268–273.

Plassman, B. L., Williams J. W., Burke, J. R., Holsinger, T., & Benjamin, S. (2010). Systematic review: Factors associated with risk for and possible prevention of cognitive decline in later life. *Annals of Internal Medicine*, 153(3), 182–193.

Project MATCH Research Group (1999). A study to remember: Response of the Project MATCH Research Group. *Addiction 94*(1), 66–69.

Quine, S., & Carter, S. (2006). Australian baby boomers' expectations and plans for their old age. *Australasian Journal on Ageing*, *25*(1), 3–8. doi:10.1111/j.1741–6612.2006.00147.x

Qureshi, I., & Fang, Y. (2010). Socialization in open source software projects: A growth mixture modeling approach. *Organizational Research Methods, 14*(1), 208–238.

Rabbitt, P. M. A., McInnes, L., Diggle, P., Holland, F., Bent, N., Abson, V., ... Horan, M. (2004). The University of Manchester longitudinal study of cognition in normal healthy old age, 1983 through 2003. *Aging Neuropsychology & Cognition*, *11*(2–3), 245–279.

Rakic P. (2002). Neurogenesis in adult primates. *Progress in Brain Research*, *138*, 3–14.

Rao, C. R. (1958). Some statistical methods for the comparison of growth curves. *Biometrics, 14*(1), 1–17.

Raudenbush, S. W., & Xiao-Feng, L. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods, 6*(4), 387–401.

Raz, N. (2009). Decline and compensation in ageing brain and cognition: Promises and constraints. *Neuropsychology Review, 19*(4): 411–414. doi:10.1007/s11065–009–9122–1

Rebok, G. W., Ball, K., Guey, L. T., Jones, R. N., Kim, H. Y., King, J. W., ..., & Willis, S. L. (2014). Ten-year effects of the advanced cognitive training for independent and vital elderly cognitive training trial on cognition and everyday functioning in older adults. *Journal of the American Geriatrics Society*, *62*(1), 16–24.

Rebok, G. W., Carlson, M. C., & Langbaum, J. B. S. (2007). Training and maintaining memory abilities in healthy older adults: Traditional and novel approaches. *Journals of Gerontology: Psychological Sciences*, *62*(Special Issue 1), 53–61.

Ree, M. J., & di Earles, J. A. (1991). Predicting training success: Not much more than g. *Personnel Psychology*, *44*(2), 321–332.

Reuter-Lorenz, P. A., & Cappel, K. A. (2008). Neurocognitive ageing and the compensation hypothesis. *Current Directions in Psychological Science, 17*(3), 177–182.

Reuter-Lorenz, P. A., & Park, D. C. (2010). Human neuroscience and the ageing mind: a new look at old problems. *The Journals of Gerontology Series B:*

*Psychological Sciences & Social Sciences, 65B*(4), 405–415. doi:10.1093/geronb/gbq035

Rey, A. (1941). "Psychological examination of traumatic encephalopathy." *Archives de Psychologic, 28,* 286–340.

Rey, A. (1964). 'Uexamen clinique en psychologic.' Paris: Presses Universitaires de France.

Reynolds, C. A., Gatz, M., & Pedersen, N. L. (2002). Individual variation for cognitive decline: Quantitative methods for describing patterns of change. *Psychology & Aging, 17*(2), 271–287. doi:10.1037/0882-7974.17.2.271

Riegel, K. F., & Riegel, R. M. (1972). Development, drop, and death. *Developmental Psychology, 6*(2), 306–319.

Rodriguez, D., Moss, H. B., & Audrain-McGovern, J. (2005). Developmental heterogeneity in adolescent depressive symptoms: Associations with smoking behaviors. *Psychosomatic Medicine, 67*(2*),* 200–210.

Roeder, K., Lynch, K. G., & Nagin, D. S. (1999). Modeling uncertainty in latent class membership: A case study in criminology. *Journal of American Statistical Association, 94*(447), 766–776.

Rosenzweig, M. R., Krech, D., Bennett, E. L., & Diamond, M. C. (1962). Effects of environmental complexity and training on brain chemistry and anatomy: A replication and extension. *Journal of Comparative & Physiological Psychology 55*(4), 429–437.

Roth, M., Tomlinson B. E., & Blessed, G. (1966). Correlation between scores for dementia and counts of 'senile plaques' in cerebral grey matter of elderly subjects. *Nature, 209*(5018), 109–110.

Rowe, J. W., & Kahn. R. L. (1987). Human ageing: Usual and successful. *Science*, *237*(4811), 143–149.

Saczynski, J. S., Margett, J. A., & Willis, S. L. (2004). Older adults' strategic behaviour: Effects of individual versus collaborative training. *Educational Gerontology, 30* (7), 587–610.

Saczynski, J. S., Rebok, G. W., Whitfield, K. E., & Plude, D. L. (2007). Spontaneous production and use of mnemonic strategies in older adults. *Experimental Aging Research*, *33*(3), 273–295.

Salthouse, T. A. (1981). Converging evidence for information-processing stages: A comparative influence stage analysis method. *Acta Psychologica*, *47*(1), 39–61.

Salthouse, T. A. (1984). Effects of age and skill in typing. *Journal of Experimental Psychology: General*, *113*(3), 345–371.

Salthouse, T. A. (1991). Cognition, motor behavior, and the assessment of atypical ageing. In M. L. Howe, M. J. Stones, & C. J. Brainerd (Eds.), *Cognitive & Behavioral Performance Factors in Atypical Aging* (pp. 425-489). New York, NY: Springer-Verlag.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403–428.

Salthouse, T. A. (2006). Mental exercise and mental ageing: Evaluating the validity of the "Use it or lose it" hypothesis. *Perspectives on Psychological Science*, *1*(1), 68–87. doi:10.1111/j.1745–6916.2006.00005.x

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, *30*(4), 507–514. doi:10.1016/j.arr.2004.05.001

Salthouse, T. A. (2012), Robust cognitive change. *Journal of International Neuropsychological Society*, *18*(4), 749–756. doi:10.1017/ S1355617712000380

Salthouse, T. A., Babcock, R. L., Skovronek, E., Mitchell, D. R. D., & Palmon, R. (1990). Age and experience effects in spatial visualization. *Developmental Psychology, 26*(1), 129–136.

Sampson, R. J., Laub, J., & Eggleston, E. P. (2004). On the robustness and validity of groups. *Journal of Quantitative Criminology, 20*, 37–42.

Sattler, J. M., & Dumont, R. (2004). *Assessment of Children: WISC-IV and WPPSI-III Supplement.* La Mesa, CA: J. M. Sattler.

Scarmeas, N., Levy, G., Tang, M.-X., Many J., & Stern, Y. (2001). Influence of leisure activity on the incidence of Alzheimer's disease. *Neurology, 57*(12), 2236–2242.

Scarmeas, N., & Stern, Y. (2003). Cognitive reserve and lifestyle. *Journal of Clinical & Experimental Neuropsychology*, *25*(5), 625–633.

Schaie, K. W. (1994). The course of adult intellectual development. *American Psychologist, 49*(4), 304–313. doi:10.1037/0003– 066X.49.4.304

Schaie, K. W. (1996). *Intellectual Development in Adulthood: The Seattle Longitudinal Study*. New York, NY: Cambridge University Press.

Schaie, K. W. (2000). The impact of longitudinal studies on understanding development in young adulthood to old age. *International Journal of Behavioural Development, 24*(3), 257–266.

Schaie, K. W. (2008). A lifespan developmental perspective of psychological ageing. In K. Laidlaw, & B. G. Knight (Eds.), *The Handbook of Emotional Disorders* *in Late Life: Assessment & Treatment* (pp. 3–32). Oxford, UK: Oxford University Press.

Schaie, K. W., & Hertzog, C. (1986). Toward a comprehensive model of adult intellectual development: Contributions of the Seattle Longitudinal Study. In R. J. Steinberg (Ed.), *Advances in Human Intelligence* (*13*, pp. 79–118). Hillsdale, NJ: Erlbaum.

Schaie, K. W., & Hofer, S. M. (2001). Longitudinal studies in research on ageing. In J. E. Birren, & K. Warner Schaie (Eds.), *Handbook of the Psychology of Aging* (5th ed.), pp. 53–77. San Diego, CA: Academic Press.

Schaie, K. W., & Willis, S. L. (1986). Can decline in adult intellectual functioning be reversed? *Developmental Psychology*, *22*(2), 223–232. doi:10.1037/0012–1649.22.2.223

Schaie, K. W., & Willis, S. L. (Eds.). (2010). *Handbook of the Psychology of Aging*. San Diego, CA: Academic Press.

Schaie, K. W, Willis, S. L., Hertzog, C, & Schulenberg, J. E. (1987). Effects of cognitive training upon primary mental ability structure. *Psychology & Aging*, *2*, 233–242.

Schaie, K., Willis, S., & Pennak, S. (2005). An historical framework for cohort differences in intelligence. *Research in Human Development*, *2*(1&2), 43–67.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177.

Scher, A. M., Young, A. C., & Meredith, W. M. (1960). Factor analysis of the electrocardiogram. *Circulation Research*, *8*(3), 519–526.

Schmidt, M. (1996). *Rey Auditory-Verbal Learning Test.* Los Angeles: Western Psychological Services.

Schneider, N. C., & Yvon, C. (2012). A review of multidomain interventions to support healthy cognitive ageing. *Journal of Nutrition, Health & Aging, 17*(3), 252–257. doi:10.1007/s12603–012–0402–8

Schooler, C. (1987). Psychological effects of complex environments during the lifespan: A review and theory. In C. Schooler, & K. W. Schai (Eds.), *Cognitive Functioning & Social Structure Over the Life Course* (pp. 22–49). Norwood, NJ: Ablex.

Schooler, C., & Mulatu, M. S. (2001). The reciprocal effects of leisure time activities and intellectual functioning in older people: A longitudinal analysis. *Psychology & Aging, 16*(3), 466–482.

Schroder, M. D., Snyder, P. J., Sielski, I., & Mayes, L. (2004). Impaired performance of children exposed in utero to cocaine on a novel test of visuospatial working memory. *Brain & Cognition, 55*, 409–412.

Schunn, C. D., & Reder, L. M. (2001). Another source of individual differences: Strategy adaptivity to changing rates of success. *Journal of Experimental Psychology: General, 130*, 59–76.

Schwartz, G., 1978. Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.

Sclove, L. S. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*, 333–343.

Shadish, W., Cook, T. D., & Campbell, D. T. (2002). *Experimental & Quasi-experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.

Sheikh, J. I., Hill, R. D., & Yesavage, J. A. (1986). Long term efficacy of cognitive training for age associated memory impairment: A six month follow-up study. *Developmental Neuropsychology, 2*, 413–421.

Shimamura, A. P., Berry, J. M., Mangels, J. A., Rusting, C. L., & Jurica, P. J. (1995). Memory and cognitive abilities in university professors: Evidence for successful ageing. *Psychological Science, 6*, 271–277.

Singer, T., Linderberger, U., & Baltes, P. B. (2003). Plasticity of memory for new learning in very old age: A story of major loss? *Psychology & Aging, 18*(2), 306–317. doi:10.1037/0882–7974.18.2.306

Singh-Manoux, A., Richards, M., & Marmot, M. (2003). Leisure activities and cognitive function in middle age: Evidence from the Whitehall II study. *Journal of Epidemiological Community Health, 57*(11), 907–913.

Sitzer, D. I., Twamley, E. W., & Jeste, D. V. (2006). Cognitive training in Alzheimer's disease: A meta-analysis of the literature. *Acta Psychiatrica Scandinavica*, *114*(2), 75–90.

Small, B. J., Dixon, R. A., McArdle, J. J., & Grimm, K. J. (2012). Do changes in lifestyle engagement moderate cognitive decline in normal ageing? Evidence from the Victoria longitudinal study. *Neuropsychology, 26*(2):144–155. doi:10.1037/a0026579

Small, G. W., Silverman, D. H. S., Siddarth, P., Ercoli. L., Miller, K. J., Lavretsky, H. … & Phelps, M. E. (2006). Effects of a 14-day healthy longevity lifestyle program on cognition and brain function. *American Journal of Geriatric Psychiatry, 14*(6), 538–545.

Snyder, P. J., Bednar. M., Cromer, J. R., & Maruff, P. (2005a). Reversal of scopolamine-induced deficits with a single dose of donezepil, an acetylcholinesterase inhibitor. *Alzheimer's & Dementia, 1*, 126–135.

Snyder, P. J., Jackson, C. E., Piskulic, D., Olver, J., Norman, T., & Maruff, P. (2008). Spatial working memory and problem solving in schizophrenia: The effect of symptom stabilization with atypical antipsychotic medication. *Psychiatry Research* 160(3):316-26. doi:10.1016/j.psychres.2007.07. 011

Snyder, P. J., Werth, J., Giordani, B., Caveney, A. F., Feltner, D., & Maruff, P. (2005b). A method for determining the magnitude of change across different cognitive functions in clinical trials: The effects of acute administration of two different doses alprazolam. *Human Psychopharmacology, 20*, 263–273.

Staff, R. T., Murray, A. D., Deary, I. J., & Whalley, L. J., (2004). What provides cerebral reserve? *Brain, 127*(5), 1191–1199.

Sterba, S. K., Prinstein, M. J., & Cox, M. J. (2007). Trajectories of internalizing problems across childhood: Heterogeneity, external validity, and gender differences. *Development & Psychopathology, 19*(2), 345–366.

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of International Neuropsychological Society, 8(3)*, 448–460.

Stern, Y. (Ed.). (2007). *Cogntive Reserve: Theory & Applications.* New York, NY: Taylor Francis.

Stern, Y. (2009). Cognitive reserve. *Neuropsychologia*, *47*(10): 2015–2028. doi:10.1016/ j.neuropsychologia.2009.03.004

Stern, Y., Gurland, B., Tatemichi, T. K., Tang, M. X., Wilder, D., & Mayeux, R. (1994). Influence of education and occupation on the incidence of Alzheimer's disease. *Journal of the American Medical Association, 271*, 1004–1010.

Stern, Y., Zarahn, E., Hilton, H. J., Flynn, J., DeLa Paz, R., & Rakitin, B. (2003). Exploring the neural basis of cognitive reserve. *Journal of Clinical & Experimental Neuropsychology, 25*(5), 691–701.

Stigsdotter-Neely, A.S., & Bäckman, L. (1989). Multifactorial memory training with older adults: How to foster maintenance of improved performance. *Gerontology*, *35*(5–6), 260–267.

Stigsdotter-Neely, A. S., & Bäckman, L. (1993). Long-term maintenance of gains from memory training in older adults: Two 3½-year follow-up studies. *Journals of Gerontology: Psychological Sciences*, *48*(5), P233–P237.

Stigsdotter-Neely, A. S., & Bäckman, L. (1995). Effects of multifactorial memory training in old age: Generalizability across tasks and individuals. *Journals of Gerontology: Psychological Sciences, 50*(3), 134–140.

Stine-Morrow, E. A. L., Miller, L. M. S., & Hertzog, C. (2006). Aging and self-regulated language processing. *Psychological Bulletin, 132*(4), 582–606.

Storandt, M. (2008). Cognitive deficits in the early stages of Alzheimer's disease. *Current Directions in Psychological Science, 17*(3), 198–202.

Strauss, E., Shermann, E. M. S., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, & Commentary* (3rd ed.). New York, NY: Oxford University Press.

Stulz, N., Gallop, R., Lutz, W., Wrenn, G. L., & Crits-Christoph, P. (2010). Examining differential effects of psychosocial treatments for cocaine dependence: An application of latent trajectory analyses. *Drug & Alcohol Dependence*, *106*(2), 164–172.

Stuss, D. T., Robertson, I. H., Craik, F. I., Levine, B., Alexander, M. P., Black, S., ... Winocur, G. (2007). Cognitive rehabilitation in the elderly: A randomized trial to evaluate a new protocol. *Journal of International Neuropsychological Society, 13*(1), 120–131.

Sullivan, A. M. (1964). *The Relation Between Intelligence and Transfer*. Unpublished doctoral dissertation. McGill University, Montreal, QC: Canada.

Summers, J. J., Wolf, A., Elder, S., Summers, M., Limbrick, K., & Tyler, M. (2010). Cognitive training in older adults. *Alzheimer's & Dementia*, *6*(Suppl. 1), S163–S164. doi:10.1016/j.jalz.2010.05.507

Teng, E. L., & Chui, H. C. (1987). The Modified Mini-Mental State (3MS) examination. *Journal of Clinical Psychiatry, 48*(8): 314–318.

Terrera, M.G. Brayne, C., Mathews, F., CC75 Study Collaboration Group. (2010). One size fits all? Why we need more sophisticated analytical methods in the explanation of trajectories of cognition in older age and their potential risk factors. International Psychogeriatics, 22(2), 291-299. doi 10.1017/ S1041610209990937

Tesky, V. A., Thiel, C., Banzer, W., & Pantel, J. (2011). Effects of a group program to increase cognitive performance through cognitively stimulating leisure activities in healthy older subjects: The AKTIVA study. *GeroPsych: Journals of Gerontopsychology & Geriatric Psychiatry, 24*(2), 83–92.

Thompson, G., & Foth, D. (2005). Cognitive-training programs for older adults: What are they and can they enhance mental fitness? *Educational Gerontology, 31*(8), 603–626.

Tierney, M. C. Nores, A. Snow, W. G. Fisher, R. H. Zorzitto, M. L. and Reid D. W. (1994). Use of the Rey Auditory Verbal Learning Test in differentiating normal ageing from Alzheimer's and Parkinson's dementia. *Psychological Assessment*, vol. 6, no. 2, pp. 129–134.

Tomlinson, B. E., Blessed, G., & Roth, M. (1976). Observations on the brains of demented old people. *Journal of Neurological Sciences, 11*(3), 205–242.

Touron, D. R., & Hertzog, C. (2004). Distinguishing age differences in knowledge, strategy use, and confidence during strategic skill acquisition. *Psychology & Aging, 19,* 452–466.

Trachtenberg, J. T., Chen, B. E., Knott, G. W., Feng, G., Sanes, J. R., Welker, E., & Svoboda, K. (2002). Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. *Nature, 420*(6917), 788–794.

Treiber, K. A., Carlson, M. C., Corcoran, C., Norton, M. C., Breitner, J. C., Piercy, K. W., ... Tschanz, J. T. (2011). Cognitive stimulation and cognitive and functional decline in Alzheimer’s disease: The Cache County Dementia Progression Study. *Journals of Gerontology: Psychological Sciences*, *66*(4), 416–425.

Troyer, A. K. (2001). Improving memory knowledge, satisfaction, and functioning via an education and intervention program for older adults. *Aging, Neuropsychology & Cognition*, *8*(4), 256–268.

Tucker, L. R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika, 23*(1), 19–23

Tucker-Drob, E. M., & Salthouse, T. A. (2009). Confirmatory factor analysis and multidimensional scaling for construct validation of cognitive abilities. *International Journal of Behavioral Development, 33,* 277–285.

Tucker-Drob, E. M., & Salthouse, T. A. (2011). Individual differences in cognitive ageing. In T. Chamorro-Premuzic, A. Furnham, & S. von Stumm (Eds.), *The Wiley-Blackwell Handbook of Individual Differences* (Section 2.9, pp. 242–268). UK: Wiley Blackwell.

Uchiyama, C.L. D'Elia, L.F. Dellinger, A.M. Becker, J.T. Selnes, O.A. Wesch, J.E. (1995). Alternate forms of the auditory-verbal learning test: Issues of test comparability, longitudinal reliability, and moderating demographic variables. *Archives of Clinical Neuropsychology, 10*, 133–145.

Uher, R., Muthén, B., Souery D., Mors, O., Jaracz, J., Placentino, A., … McGuffin, P. (2010). Trajectories of change in depression severity during treatment with antidepressants. *Psychological Medicine, 40*(8). 1367–1377.

United Nations. (2009). *World Population Prospects: The 20 08 Revision, Vol. I, Comprehensive Tables.* United Nations publication, Department of Economic and Social Affairs, Population Division. Sales No. E.07.XIII.2.

Uttl, B. (2005). Measurement of individual differences lessons from memory assessment in research and clinical practice. *Psychological Science*, *16*(6), 460–467.

Vakil, E., Greenstein, Y., & Blachstein, H. (2010). Normative data for composite scores for children and adults derived from the Rey auditory verbal learning test. *Clinical Neuropsychologist, 24*(4), 662–677. doi:10.1080/13854040903493522

Valenzuela, M. J. (2008). Brain reserve and the prevention of dementia. *Current Opinion in Psychiatry, 21*(3), 296–302.

Valenzuela, M. J., Leon, I., Suo, C., Piamba, D. M., Kochan, N., Brodaty, H., & Sachdev, P. (2013). Cognitive lifestyle in older persons: The population-based

Sydney memory and ageing study. *Journal of Alzheimer's Disease, 36*(1), 87–97.

Valenzuela, M., & Sachdev, P. (2006a). Brain reserve and cognitive decline: A non-parametric systematic review. *Psychological Medicine, 36*(8), 1065–1073.

Valenzuela, M., & Sachdev, P. (2006b). Brain reserve and dementia: A systematic review. *Psychological Medicine, 36*(4), 441–454.

Valenzuela, M., & Sachdev, P. (2009). Can cognitive exercise prevent the onset of dementia? Systematic review of randomized clinical trials with longitudinal follow-up. *The American Journal of Geriatric Psychiatry*, *17*(3), 179–187.

van Boxtel, M. P., Langerak, K., Houx, P. J., & Jolles, J. (1996). Self-reported physical activity, subjective health, and cognitive performance in older adults. *Experimental Aging Research*, *22*(4), 363–379.

van den Burg, W., & Kingma, A. (1999). Performance of 225 Dutch school children on Rey's Auditory Verbal Learning Test (AVLT): Parallel test–retest reliabilities with an interval of 3 months and normative data. *Archives of Clinical Neuropsychology*, *14*, 545–559.

van Hooren, S. A. H., Valentijn, A. M., Bosma, H., Ponds, R. W. H. M., Van Boxtel, M. P. J., & Jolles, J. (2007). Cognitive functioning in healthy older adults aged 64–81: A cohort study into the effects of age, sex, and education. *Aging, Neuropsychology, & Cognition*, *14*(1), 40–54..doi:10.1080/138255890969483

Verghese, J., LeValley, A., Derby, C., Kuslanksy, G., Katz, M., Hall, C., … Lipton, R. B. (2006). Leisure activities and the risk of amnestic mild cognitive impairment in the elderly. *Neurology*, *66*(6), 821–827.

Verghese, J., Lipton, R. B., Katz, M. J., Hall, C. B., Derby, C. A., Kuslansky, G., ... Buschke, H. (2003). Leisure activities and the risk of dementia in the elderly.

*New England Journal of Medicine, 348*(25), 2508–2516. doi:10.1056/NEJMoa022252

Verhaeghen, P., & Marcoen, A. (1996). On the mechanism of plasticity in young and older adults after instruction on the Method of Loci: Evidence for an amplification model. *Psychology & Aging, 11*(1), 164–178.

Verhaeghen, P., Marcoen, A., & Goossens, L. (1992). Improving memory performance in the aged through mnemonic training: A meta-analytic study. *Psychology & Aging, 7*(2), 242–251.

Verhaeghen, P., & Salthouse, T. A. (1997). Meta-analyses of age–cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin*, *122*(3), 231.

Wahlin, Å., Backman, L., Mantyla, T., Herlitz, A., Viitanen, M., & Winblad, B. (1993). Prior knowledge and face recognition in a community-based sample of healthy, very old adults. *Journals of Gerontology: Psychological Sciences, 48*(2), 54–61.

Wang, J. Y. J., Zhou, D. H. D., Li, J., Zhang, M., Deng, J., Tang, M., ... Chen, M. (2006). Leisure activity and risk of cognitive impairment: The Chongqing ageing study. *Neurology, 66*(6), 911–913.

Wang, M. (2007). Profiling retirees in the retirement transition and adjustment process: Examining the longitudinal change patterns of retirees' psychological well-being. *Journal of Applied Psychology, 92*(2), 455–474.

Webber, J., Schermann, B., McCall, C., & Coleman, M. (1993). Research on self-monitoring as a behavior management technique in special education classrooms: A descriptive review. *Remedial & Special Education, 14*(2), 38–56.

Wechsler, D. (2001). *Wechsler Test of Adult Reading*. San Antonio, TX: The Psychological Corporation.

Weil, A., & Small, G. (2007). *The Healthy Brain Kit.* Boulder, CO: Sounds True, Inc.

West, R. L., Bagwell, D. K., & Dark-Freudeman, A. (2008). Self-efficacy and memory ageing: The impact of a memory intervention based on self-efficacy. *Aging Neuropsychology & Cognition, 15*(3), 302–329.

West, R. L., Crook, T. H., & Barron, K. L. (1992). Everyday memory performance across the life span: Effects of age and noncognitive individual differences. *Psychology & Aging, 7*(1), 72–82.

West, R. L., & Hastings, E. C. (2011). Self-regulation and recall: Growth curve modeling of intervention outcomes for older adults. *Psychology & Aging, 26*(4), 803–812. doi:10.1037/a0023784

West, R. L., & Tomer, A. (1989). Everyday memory problems of healthy older adults: Characteristics of a successful intervention. In G. C. Gilmore, P. J. Whitehouse, & M. L. Wykle (Eds.), *Memory, Aging, & Dementia: Theory, Assessment, & Treatment* (pp. 75–98). New York, NY: Springer.

Westerman, R., Darby, D. G., Maruff, P., & Collie, A. (2001). Computer-assisted cognitive function assessment of pilots. *ADF Health – Journal of the Australian Defence Health Service, 2*, 29–36.

Wexley, K. N. (1984). Personnel training. *Annual Review of Psychology*, *35*(1), 519–551.

Whalley, L. J., Deary, I. J., Appleton, C. L., & Starr, J. M. (2004). Cognitive reserve and the neurobiology of cognitive ageing. *Ageing Research Reviews, 3*(4), 369–382. doi:10.1016/j.arr.2004.05.001

Whitlock, L. A., McLaughlin, A. C., & Allaire, J. C. (2012). Individual differences in response to cognitive training: Using a multi-modal, attentionally demanding game-based intervention for older adults. *Computers in Human Behavior, 28*(4), 1091–1096.

Wiederholt, W. C., Cahn, D., Butters, N. M., Salmon, D. P., Kritz- Silverstein, D., & Barrett-Connor, E. (1993). Effects of age, gender and education on selected neuropsychological tests in an elderly community cohort. *Journal of American Geriatrics Society, 41*, 639– 647.

Willis, S.L. (1989). Cohort differences in cognitive ageing: A sample case. In K.W. Schaie & C. Schooler (Eds.), *Social Structure and Aging: Psychological Processes* (pp. 94–112). Hillsdale, NJ: Erlbaum.

Willis, S. L., & Schaie, K. W. (1987). *Annual Review of Gerontology and Geriatrics.* New York: Springer.

Willis, S. L., & Nesselroade, C. S. (1990). Long-term effects of fluid ability training in old-old age. *Developmental Psychology, 26*(6), 905–910. doi:10.1037/ 0012–1649.26.6.905

Willis, S. L., & Schaie, K. W. (2009). Cognitive training and plasticity: theoretical perspective and methodological consequences. *Restorative Neurology & Neuroscience, 27*(5), 375–389.

Willis, S. L., Schaie, K. W., & Martin, M. (2009). Cognitive plasticity. In V. L. Bengtson, D. Gans, N. M. Puntney, & M. Silverstein (Eds.), *Handbook of Theories of Aging* (2nd ed.), pp. 295–322. New York, NY, Springer.

Willis, S. L., Tennstedt, S. L., Marsiske, M., Ball, K., Elias, J., Koepke, K. M., ... Wright, E. (2006). Long-term effects of cognitive training on everyday functional outcomes in older adults. *Journal of American Medical Association*, *296*(23), 2805–2814.

Wilson, R. S., Hebert, L. E., Scherr, P. A., Barnes, L. L., De Leon, C. M., & Evans, D. A. (2009). Educational attainment and cognitive decline in old age. *Neurology*, *72*(5), 460–465.

Wilson, R. S., Beck, T. L., Bienias, J. L., & Bennett, D. A. (2007). Terminal cognitive decline: Accelerated loss of cognition in the last years of life. *Psychosomatic Medicine, 69*(2), 131–137.

Wilson, R. S., Beckett, L. A., Barnes, L. L., Schneider, J. A., Bach, J., Evans, D. A., & Bennett, D. A. (2002a). Individual differences in rates of change in cognitive abilities of older persons. *Psychology & Aging*, *17*(2), 179–193.

Wilson, R. S., Leurgans, S. E., Bole, P. A., Schneider, J. A., & Bennett, D. A. (2010). Neurodegenerative basis of age-related cognitive decline. *Neurology, 75*(12), 1070–1078.

Wilson, R. S., Mendes de Leon, C. F., Barnes, L. L., Schneider, J. A., Bienias, J. L., & Evans, D. A. (2002b). Participation in cognitively stimulating activities and risk of incident Alzheimer disease. *Journal of the American Medical Association, 287*, 742–748.

Winocur, G., Craik, F. I., Levine, B., Robertson, I. H., Binns, M. A., Alexander, M., ... Stuss, D. T. (2007a). Cognitive rehabilitation in the elderly: Overview and future directions. *Journal of International Neuropsychological Society*, *13*(1), 166–171.

Winocur, G., Palmer H., Dawson, D., Binns, M. A., Bridges, K., & Stuss, D. T. (2007b). Cognitive rehabilitation in the elderly: An evaluation of psychosocial factors. *Journal of International Neuropsychological Society*, *13*(1), 153–165.

Woolverton, M., Scogin, F., Shackleford, S. B., & Duke, L. (2001). Problem-targeted memory training for older adults. *Ageing, Neuropsychology & Cognition*, *8*(4), 241–255.

World Health Organization. (2002). *Active Ageing – A Policy Framework*. Geneva, Switzerland: World Health Organization. www.who.int/hpr/ageing/ publications.htm

Wu, J., & Witkiewitz, K. (2008.). Network support for drinking: An application of multiple groups growth mixture modeling to examine client-treatment matching. *Journal of Studies on Alcohol & Drugs*, *69*, 5.

Yaffe, K., Fiocco, A. J., Lindquist, K., Vittinghoff, E., Simonsick, E. M., Newman, A. B., ... Harris, T. B. (2009). Predictors of maintaining cognitive function in older adults: The Health ABC Study. *Neurology*, *72*(23), 2029–2035. doi:10.1212/WNL. 0b013e3181a92c36

Yaffe, K., Lui, L.-Y., Zmuda, J., & Cauley, J. (2002). Sex hormones and cognitive function in older men. *Journal of American Geriatrics Society*, *50*(4), 707–712. doi:10.1046/j.1532–5415.2002.50166.x

Yesavage, J. A., Sheikh, J. I., Friedman, L., & Tanke, E. D. (1990). Learning mnemonics: Roles of ageing and subtle cognitive impairment. *Psychology & Aging, 5,* 133–137.

Yesavage, J. A., Sheikh, J., Tanke, E. D., & Hill, R. (1988). Response to memory training and individual differences in verbal intelligence and state anxiety. *American Journal of Psychiatry, 145*(5), 636–639.

Yevchak, A. M., Loeb, S. J., & Fick, D. M. (2007). Promoting cognitive health and vitality: A review of clinical implications. *Geriatric Nursing*, *29*(5), 302–310. doi:10.1016/j.gerinurse.2007.10.017

Zelinski, E. M. (2009). Far transfer in cognitive training of older adults. *Restorative Neurology & Neuroscience*, *27*(5), 455–471.

Zelinski, E. M., Gilewski, M. J., & Schaie, K. W. (1993). Individual differences in cross-sectional and three-year longitudinal memory performance across the adult lifespan. *Psychology & Aging*, *8*(2), 176–186.

Zelinski, E., Housen, P., Yaffe, K., Ruff, R., Kennison, R., Mahncke, H. E., & Smith, G. E. (2008). Who benefits from computer training of cognitive abilities? *Gerontechnology 7*(2), 250.

Zelinski, E. M., Spina, L. M., Yaffe, K., Ruff, R., Kennison, R. F., Mahncke, H. W., and Smith, G. E. (2011). Improvement in memory with plasticity-based adaptive cognitive training: results of the 3-month follow-up. *Journal of the American Geriatriatic Society* 59, 258–265.

Zelinski, E. M., & Stewart, S. (1998). Individual differences in 16-year memory changes. *Psychology & Aging*, *13*(4), 622–630.

Zelinski, E. M., Yaffe, K., Ruff, R., Kennison, R., Smith, G. (2007). *The IMPACT Study: A Randomized Controlled Trial of a Brain Plasticity-Based Training Program for Age-Related Cognitive Decline.* Presented at the annual meeting of the Gerontological Society of America, San Francisco, CA.

Zwaagstra, R., Schmidt, I., & Vanier, M. (1996). Recovery of speed of information processing in closed head injury patients. *Journal of Clinical & Experimental Neuropsychology*, *18*, 383–393.

# APPENDIX 1

## Components of Active Cognitive Enhancement Training Sessions

### Educational Lecturettes

The educational lecturettes provided psychoeducation in relation to factors influencing cognitive loss in ageing. They also had a risk reduction emphasis, highlighting lifestyle factors believed to maximise cognitive health. Weekly topics were titled:

1. ‘How Memory Works’ (*e.g.*, teaching participants about encoding, consolidation retrieval, normalising forgetting)
2. ‘What is dementia?’ (including information on risk factors and differences with normal ageing)
3. ‘Mental Fitness (*Use it or lose it*)’
4. ‘Physical Fitness’
5. ‘Diet and Nutrition’ (which was developed in consultation with a dietician/nutritionist and emphasised a Mediterranean diet)
6. ‘Lifestyle factors’ (including information about the risks of drug use such as smoking)
7. ‘Health Checks’ (information about blood pressure, blood glucose, and cholesterol)
8. ‘Stress Management’
9. ‘Avoiding Head Injury’
10. ‘Keeping an Active Social Life’.

The lecturette material was adapted from information from the ‘Mind Your Mind’ program: a psychoeducation program conducted at AATas and expanded using current literature (*e.g.*, Scarmeas et al., 2001, Einstein & McDaniel, 2004).

Lecturettes were presented as PowerPoint presentations and facilitators provided opportunities for participants to discuss the topic and ask questions. This was component included because there is increasing evidence that the most effective training programs incorporate educational material about the process of ageing and lifestyle activities (Hohaus, 2007; Small et al., 2006; Stuss et al., 2007). Interactive groups have also been shown to increase efficacy of training programs (Verhaeghen et al., 1992).

### Physical Activity

Physical exercise was of low intensity and including walking for approximately 5-10 minutes and/or a basic stretches designed by a physiotherapist to improve balance and coordination. The exercise was designed with those who may have some physical limitations in mind and participants were instructed to exercise within their own limits (*i.e.*, to stop if they experienced pain). This component was designed to demonstrate to participants the simplicity of exercise and encourage them to begin or increase the amount of exercise in which they engage. An additional aim was to give participants a small break after sitting and listening.

### Mnemonics

Memory strategies were considered the central component of the program. Strategies taught included: visual imagery; tips for remembering names; story making;

categorisation; chunking; method of loci; association; rote and repetition; acronyms and acrostics; and use of external memory aids.

The mnemonics were selected given that they can be applied to various materials and learning conditions (Belleville et al., 2006) and that they had been frequently used in cognitive training and rehabilitation groups (Belleville et al., 2006; Fairchild et al., 2012; Gatz, 2005; Langbaum et al., 2009).

### Word Memory Task

Participants were presented with a 20–30 nouns projected at the front of the room for approximately 1 minute and were asked to apply one of the memory strategies to which they had been introduced. Novel words were presented each session. Participants timed their performance using a stopwatch.

### Refreshment Break

A refreshment break was included to provide participants with a mental break. It was also used as a time to socialise, in light of research showing that social interactions may foster the maintenance of cognitive function in later life and enhance adherence to intervention programs (Fratiglioni et al., 2004; Verghese et al., 2003; Verhaeghen et al., 1992). Facilitators also encouraged social interaction between group members outside of the program.

### Visual Memory Exercise

This task required participants to observe items for approximately 2 minutes and then recall objects presented on the screen. New stimuli were used each session and participants timed their performance using a stopwatch.

### Arithmetic Exercise

This exercise was a calculation task involving addition, subtraction and multiplication problems (no division). Participants were asked to work as quickly (and as accurately) as they could and timed their performance using a stopwatch. The task was designed to enhance speed of processing (Weil & Small, 2007).

### Relaxation/meditation

This element of the program included facilitator-guided slow breathing, visualisation and release-only relaxation (an adaptation of progressive muscle relaxation better suited to older adults where there are pain issues, such as arthritis), techniques which were incorporated for stress management. Mindfulness meditation techniques were also executed for both stress reduction and due to its purported benefits in the development of focussed attention (Belleville et al., 2006; Stigsdotter-Neely & Bäckman, 1995).

### Group Activities

Participants were asked to work together in small teams (3-6 members, most often consisting of participants from the same table) to solve different, timed word puzzles or 'brain teasers'. Research has demonstrated that older adults benefit from collaborating with others on varied cognitive tasks (Saczynski, Margett, & Willis, 2004).

### Homework

Activities involved the application of the learned memory strategies to solving everyday problems. Participants were also asked to perform an unusual activity (*e.g.*, brushing teeth with non-dominant hand) to stimulate brain areas not normally used for the activity (Katz & Rubin, 1999). Homework performance was discussed and reviewed at the start of the following session and homework sheets were collected by facilitators. A group discussion review of homework was conducted to stimulate participant self-reflection on their performance.

Research suggests that self-monitoring can be particularly instrumental to increase a desired behaviour (Webber, Schermann, McCall, & Coleman, 1993). Homework was also incorporated to reinforce learning that had taken place during sessions and to assist with the application of content learnt to participants' day-to-day lives. It also aimed to encourage participants to try novel activities, which is considered a key aspect of cognitive stimulation (and motivation). Homework exercises are considered to offer practice of strategies in more ecological and diverse situations, and to develop expertise (Belleville et al., 2006).

# APPENDIX 2

## Experimental Participant Consent Form

UTAS
FACULTY OF SCIENCE, ENGINEERING & TECHNOLOGY
School of Psychology

## CONSENT FORM

Title of Project: **Evaluation of a Multidimensional Cognitive Enhancement Programme for Healthy Older Adults**

---

1. I have read and understood the 'Information Sheet' for this project.
2. The nature and possible effects of the study have been explained to me.
3. I understand that the study involves undertaking neuropsychological tests and questionnaires about memory, thinking and feelings in four sessions of 1 hour each (totalling 4 hours) and the A.C.E training programme which will consist of a 1.5 to 2 hour session conducted once a week, over 10 weeks.
4. I understand that I may be video-recorded during both the ACE program and the neuropsychological tests.
5. I understand that participation involves the possibility that the researchers may detect a decline in my thinking and memory.
6. I would like to be told if a decline is detected. YES □ NO □
7. If I ticked yes: I would like to be contacted by a counsellor from Alzheimer's Australia. YES □ NO □
8. I may also experience stress or anxiety from the challenges of the testing. While this is expected to be minimal, if this occurs, the facilitator will offer me support or alternatively, arrange for me to see a counsellor.
9. I understand that all research data will be securely stored on the University of Tasmania premises for five years [or at least five years], and will then be destroyed [or will be destroyed when no longer required].
10. Any questions that I have asked have been answered to my satisfaction.
11. I agree that research data gathered from me for the study may be published provided that I cannot be identified as a participant.
12. I understand that the researchers will maintain my identity confidential and that any information I supply to the researcher(s) will be used only for the purposes of the research.

13. I agree to participate in this investigation and understand that I may withdraw at any time without any effect, and if I so wish, may request that any data I have supplied to date be withdrawn from the research.

Name of Participant:

Signature: Date:

**Statement by Investigator**

☐ I have explained the project & the implications of participation in it to this volunteer and I believe that the consent is informed and that he/she understands the implications of participation

If the Investigator has not had an opportunity to talk to participants prior to them participating, the following must be ticked.

☐ The participant has received the Information Sheet where my details have been provided so participants have the opportunity to contact me prior to consenting to participate in this project.

Name of Investigator:

Signature: Date:

# APPENDIX 3

## Control Participant Consent Form

## CONSENT FORM

Title of Project: **Evaluation of a Multidimensional Cognitive Enhancement Program for Healthy Older Adults**

1. I have read and understood the 'Information Sheet' for this project.
2. The nature and possible effects of the study have been explained to me.
3. I understand that the study involves undertaking neuropsychological tests and questionnaires about memory, thinking and feelings in four sessions of 1 hour each (totalling 4 hours).
4. I understand that I may be video-recorded during the neuropsychological tests.
5. I understand that participation involves the possibility that the researchers may detect a decline in my thinking and memory.
6. I would like to be told if a decline is detected. YES ☐ NO ☐
7. If I ticked yes: I would like to be contacted by a counsellor from Alzheimer's Australia. YES ☐ NO ☐
8. I understand that I may also experience stress or anxiety from the challenges of the testing. While this is expected to be minimal, if this occurs, the facilitator will offer me support or alternatively, arrange for me to see a counsellor.
9. I understand that all research data will be securely stored on the University of Tasmania premises for at least five years, and will be destroyed when no longer required.
10. Any questions that I have asked have been answered to my satisfaction.
11. I agree that research data gathered from me for the study may be published provided that I cannot be identified as a participant.
12. I understand that the researchers will maintain confidentiality of my identity and that any information I supply to the researcher(s) will be used only for the purposes of the research.

13. I agree to participate in this investigation and understand that I may withdraw at any time without any effect, and if I so wish, may request that any data I have supplied to date be withdrawn from the research.

Name of Participant:

Signature: Date:

**Statement by Investigator**

☐ I have explained the project & the implications of participation in it to this volunteer and I believe that the consent is informed and that he/she understands the implications of participation

If the Investigator has not had an opportunity to talk to participants prior to them participating, the following must be ticked.

☐ The participant has received the Information Sheet where my details have been provided so participants have the opportunity to contact me prior to consenting to participate in this project.

Name of Investigator:

Signature: Date:

# APPENDIX 4

## Experimental Participant Information Sheet Form

UTAS

FACULTY OF SCIENCE, ENGINEERING & TECHNOLOGY

School of Psychology

# PARTICIPANT INFORMATION SHEET

# SOCIAL SCIENCE/ HUMANITITES RESEARCH

**Evaluation of a Multidimensional Cognitive Enhancement Program for Healthy Older Adults**

---

**Invitation**

You are invited to participate in a research study into a new memory and thinking improvement program called Active Cognitive Enhancement (ACE). The program has been developed by Alzheimer's Australia (Hobart, Tasmania) and the Department of Health and Human Services (DHHS). The study is being conducted by
Professor Jeff Summers (School of Psychology, University of Tasmania),
Dr Mathew Summers (School of Psychology, University of Tasmania),
Professor James Vickers (Wicking Dementia Research and Education Centre, UTAS)
Ms Anna Wolf (School of Psychology, University of Tasmania),
Ms Kelly Limbrick (School of Psychology, University of Tasmania)
Dr Sarah Elder (Alzheimer's Australia Tasmania, Hobart), and
Mr Malcolm Tyler (Department of Health and Human Services).

**1. What is the purpose of this study?**

The purpose is to investigate whether the ACE program can produce memory and thinking improvements measured using neuropsychological tests and personal reports within healthy, community-dwelling older adults.

**2. Why have I been invited to participate in this study?**

You are eligible to participate in this study because you are:
- aged 55 years or over
- willing and able to commit to the requirements of the study
- currently living in the community
- physically and psychologically healthy
- not suffering any significant sensory impairment
- willing to use a computer
- a first time participant in a program which trains and tests memory and thinking
- able to converse in English sufficiently to understand the study requirements and provide informed consent

**4. What does this study involve?**

In groups of 15-20, you will be asked to undertake the ACE program, which will consist of a 2½ hour session conducted once a week, over 10 weeks. The program will involve individual and group activities and homework out of class, including memory tasks, word games, arithmetic, use of memory strategies, exercises to focus attention, creative thinking, physical stretching and relaxation exercises. During the sessions, educational lecturettes will be presented, which will provide you with information about normal ageing and healthy ageing, and ways to reduce your risk of dementia. Some of these sessions may be video recorded.

Before and after participating in the ACE program, you will be asked to come to two testing sessions. Again, some of these sessions may be video recorded. In the first session, you will be asked to:

- undertake an assessment of your thinking and memory
- answer questions assessing how much a given statement applies to you over the past week, for example “I found it hard to wind down”
- recall and recite a number of words within a time limit
- read and say a series of unfamiliar and uncommon words
- complete a questionnaire about your feelings about your memory, your memory mistakes, and the memory strategies you use
- complete a series of tasks on a computer that measure cognitive functions. Recording your answers will be simple and will require you to either use two buttons on the keyboard or mouse clicks.

It is expected that this first session will take around 2 hours. It is important that you understand that one of the tests administered in this initial testing session is a screening test for dementia, and therefore has the capacity to detect impaired cognitive functioning. In signing the consent form, you will be asked to nominate whether you would like to be told if impairment is in fact detected. In the event that an impairment is detected, we would speak to you in person to explain your assessment results, and advise that you see your GP for referral to a specialist for further investigation. We would seek your permission to forward your assessment results on to your GP.

In the second session you will be asked to:

- answer questions on some tasks most older adults have to do in their daily life, such as taking one’s medications, using the telephone, and using money to pay for things.
- complete a series of tasks on a computer that measure cognitive functions using either two buttons on the keyboard or mouse clicks to record your answers. This second computerised testing session will run with up to 5 other participants (although your results will not be shared with the others).
- complete a series of tasks that measure neuromotor function. The neuromotor tasks will measure balance and postural sway, finger tapping, hand and finger dexterity.

It is expected that the second session will take around 2 hours. Therefore, overall, the pre-ACE and post-ACE testing sessions will involve a total of approximately 8 hours of assessment.

We will also be investigating whether genetic factors that modify individual risk for dementia at advanced age may also play a role in the effectiveness of the ACE program in potentially improving cognitive function. On the consent form, we will ask about your willingness to provide a blood sample by venepuncture (10-15 mls) or have a buccal tissue sample taken from inside your cheek by a swab by a registered nurse at a time which is convenient for you. As the genetic information is only useful in a research context relative to the proposed study, and of no value clinically, the details of your results will be de-identified and not available to be released to participants.

It is important that you understand that your involvement is this study is voluntary. While we would be pleased to have you participate, we respect your right to decline. There will be no consequences to you if you decide not to participate, and this will not affect your treatment/service. If you decide to discontinue participation at any time, you may do so without providing an explanation. All information will be treated in a confidential manner, and your name will not be used in any publication arising out of the research. All of the research will be kept in a locked cabinet in the office of Prof. Jeffery Summers at the University of Tasmania, Sandy Bay campus.

**5. Are there any possible benefits from participation in this study?**

It is possible that you will notice an improvement in your memory and thinking from the program after a certain period of time. This may lead to an improvement in your day-to-day life. It may also result in improved confidence and lessened anxiety about your memory. We will be interested to see if you experience any other benefits from the ACE program.

The findings of this study may provide valuable information for others and it may lead to a better understanding of the ageing brain and the benefit of memory training programs.

**6. Are there any possible risks from participation in this study?**

There are no specific risks anticipated with participation in this study. However, if you find that you are becoming distressed or anxious about the results of your neuropsychological performance or involvement within the ACE program itself you will be advised to receive support from Alzheimer's Australia Tasmania or alternatively, we will arrange for you to see a counsellor at no expense to you.

**7. What if I have questions about this research?**

If you would like to discuss any aspect of this study please feel free to contact either Prof. Jeffery Summers on (03) 6226 2884 or Dr Sarah Elder on (03) 6224 3077. We would be happy to discuss any aspect of the research with you. Once we have analysed the information we will be mailing/emailing you a summary of our findings. You are welcome to contact us at that time to discuss any issue relating to the research study.

This study has been approved by the Tasmanian Social Science Human Research Ethics Committee. If you have concerns or complaints about the conduct of this study should contact the Executive Officer of the HREC (Tasmania) Network on (03) 6226 7479 or email human.ethics@utas.edu.au. The Executive Officer is the person nominated to receive complaints from research participants. You will need to quote HREC project number H 10127.

**Thank you for taking the time to consider this study.**
**This information sheet is for you to keep.**

# APPENDIX 5

## Control Participant Information Sheet Form

UTAS
FACULTY OF SCIENCE, ENGINEERING & TECHNOLOGY
School of Psychology

# CONTROL GROUP PARTICIPANT INFORMATION SHEET

# SOCIAL SCIENCE/HUMANITIES RESEARCH

### Evaluation of a Multidimensional Cognitive Enhancement Program for Healthy Older Adults

---

**Invitation**

You are invited to participate as part of a control group in a research study into new memory and thinking improvement program called Active Cognitive Enhancement (ACE). The program has been developed by Alzheimer's Australia (Hobart, Tasmania) and the Department of Health and Human Services (DHHS). The study is being conducted by

Professor Jeff Summers (School of Psychology, University of Tasmania),
Dr Mathew Summers (School of Psychology, University of Tasmania),
Professor James Vickers (Wicking Dementia Research and Education Centre, UTAS),
Ms Anna Wolf (School of Psychology, University of Tasmania),
Ms Kelly Limbrick (School of Psychology, University of Tasmania),
Dr Sarah Elder (Alzheimer's Australia Tasmania, Hobart), and
Mr Malcolm Tyler (Department of Health and Human Services).

**1. What is the purpose of this study?**

The purpose is to investigate whether the ACE program can produce memory and thinking improvements measured using neuropsychological tests and personal reports within healthy, community-dwelling older adults.

**2. Why have I been invited to participate in this study?**

You are eligible to participate in this study because you are:

- Aged 55 years or older
- Willing and able to commit to the requirements of the study
- Currently living in the community
- Sufficiently physically and psychologically healthy
- Not suffering any significant sensory impairment
- Willing to use a computer

- A first time participant in a program which trains and tests memory and thinking
- Able to converse in English sufficiently to understand the study requirements and provide informed consent
- Willing to accept that the monitoring process may detect problems with your memory or thinking.

**3. What does the study (control group) involve?**

You will be asked to come to two neuropsychological testing sessions during April-May and again during July-August 2010. You will also be asked if you are willing to have an EEG to monitor brain activity during performance of 3 cognitive tasks. (The EEG dates are as yet to be finalised). Some of the testing sessions may be video recorded.

In the first session, you will be asked to undertake an assessment of your thinking and memory

- Recall and recite a number of words over a number of trials
- Read and say a series of words, some of which may be unfamiliar or uncommon
- Complete a number of questionnaires relating to:
    - Your memory
    - How you have been feeling
    - Your current activities and medications you currently take
    - Your functioning in routine activities around the home and in the community
    - Your mental and physical activity patterns
    - Your satisfaction with and feelings about different areas of your life

You will be asked to complete a series of tasks on a computer that measure cognitive functions such as memory, learning, thinking speed and discrimination. Recording your answers will be simple and will require you to use either two buttons on the keyboard or mouse clicks. Instruction on the computer program will be given during the first session. You will also be asked to complete a series of neuromotor tests. The neuromotor tests will measure balance and postural sway, finger tapping, hand and finger dexterity.

It is expected that this first session will take up to 3 hours. It is important that you understand that one of the tests administered in this initial testing session is a screening test for dementia, and therefore has the capacity to detect impaired cognitive functioning. In signing the consent form, you will be asked to nominate whether you would like to be told if impairment is in fact detected. In the event that impairment is detected, we would speak to you in person to explain your assessment results, and advise that you see your GP for referral to a specialist for further investigation. We would seek your permission to forward your assessment results on to your GP.

In the second session you will be asked to:

- Complete a series of tasks on a computer that measure cognitive functions using either two buttons on the keyboard or mouse clicks to record your answers. This second computerised testing session will run with up to 5 other participants (although your results will not be shared with others).

It is expected that the second session will take up to 1 hour.

The EEG will be scheduled as a separate session and will take 2 hours.

Therefore, overall the testing sessions will take in total approximately 12 hours.

You are also invited to participate in long-term follow up of participants, which will take the form of six- monthly neuropsychological testing sessions for two years.

We will also be investigating whether genetic factors that modify individual risk for dementia at advanced age may also play a role in the effectiveness of the ACE program in potentially improving cognitive function. On the consent form, we will ask about your willingness to provide a blood sample by venepuncture (10-15mls) which will be obtained by a registered nurse at a time which is convenient to you. As the genetic information is only useful in a research context relative to the proposed study, and of no value clinically, the details of your results will be de-identified and not available to be released to participants.

It is important that you understand that your involvement in this study is voluntary. While we would be pleased to have you participate, we respect your right to decline. There will be no consequences to you if you decide not to participate, and this will not affect your treatment/service. If you decide to discontinue participation at any time, you may do so without providing an explanation. All information will be treated in a confidential manner, and your name will not be used in any publication arising out of the research. All of the research will be kept in a locked cabinet in the office of Prof. Jeff Summers at the University of Tasmania, Sandy Bay campus.

**4. Are there any possible benefits from participation in this study?**

The findings of this study may provide valuable information for others and it may lead to a better understanding of the ageing brain and the benefit of memory training programs.

**5. Are there any possible risks from participation in this study?**

There are no specific risks anticipated with participation in this study. However, if you find that you are becoming distressed or anxious about the results of your neuropsychological performance you will be advised to receive support from Alzheimer's Australia Tasmania or alternatively, we will arrange for you to see a counsellor at no expense to you.

**6. What if I have questions about this research?**

If you would like to discuss any aspect of this study please feel free to contact either Prof. Jeffery Summers on (03) 6226 2884 (BH) or Dr Sarah Elder on (03) 6224 3077 (BH) We would be happy to discuss any aspect of the research with you. Once we have analysed the information we will be mailing/emailing you a summary of our findings. You are welcome to contact us at that time to discuss any issue relating to the research study.

This study has been approved by the Tasmania Social Science Human Research Ethics Committee. If you have concerns or complaints about the conduct of this study you should contact the Executive Officer of the HREC (Tasmania) Network on (03) 6226 n7479 or email human.ethics@utas.edu.au. The Executive Officer is the person nominated to receive complaints from research participants. You will need to quote HREC project number H 10127.

Thank you for taking the time to consider this study.

This information sheet is for you to keep.

# APPENDIX 6

# Control and Treatment Groups' Model Results for Verbal Memory

## Control Group Verbal Memory Performance

### Latent Growth Modelling of the Verbal Memory Scores for the Control Group

Table A.1 presents the model fit indices of the 1- to 4-class growth models (the LGM and the LCGA) to determine the optimal model of VM performance for the control group. As noted in the data analysis section of the Method chapter, MPlus parameter default options were used. As can be seen in the table, the LGM (the single-class model) in which variances were fixed (*e.g.*, Nagin, 1999) for the control group resulted in a BIC = 958.549 and an ABIC = 939.671. The VM trajectory is identified in Figure A.1. Whilst there was an increase in performance scores across the 12-month period, controls did not demonstrate significant growth in VM performance ($n$ = 62, estimate = 0.043, $SE$ = 0.028, $p$ = .13). A LGM in which variances were allowed to vary (*e.g.*, Muthén & Muthén, 2000) produced an inadmissible model, due to negative variance for the slope parameter. As such, as previously highlighted, subsequent control models were conducted with variances fixed.

Table A.1.
*Model Fit Indices from the Verbal Memory (VM) Latent Growth Modelling (LGM) for the Control Group*

| Model | BIC | ABIC | LMR *p* | Adjusted LRT *p* | Entropy | Class membership (%) C1 | C2 | C3 | C4 |
|---|---|---|---|---|---|---|---|---|---|
| 1-class | 958.549 | 939.671 | | | | 100 | | | |
| 2-class | 907.892 | 879.576 | .16 | .17 | 0.845 | 22.21 | 77.79 | | |
| **3-class** | **884.532** | **846.777** | **.0002*** | **.0003*** | **0.869** | **3.56** | **30.58** | **65.86** | |
| 4-class | 896.914 | 849.719 | .50 | .50 | 0.896 | 0.00 | 3.56 | 30.5 | 65.86 |

*Note:* Bold indicates best fit. BIC = Bayesian information criterion; ABIC = sample-size adjusted Bayesian Information Criterion; LMR = Vuong-Lo-Mendell-Rubin likelihood ratio test; Adjusted LRT = Lo-Mendell-Rubin Adjusted likelihood ratio test. * $p$ < .001, two-tailed.

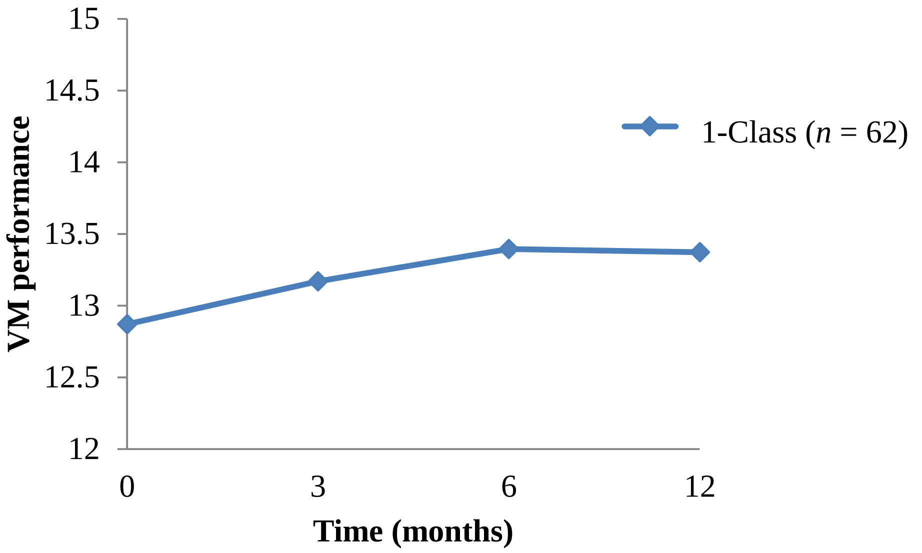


*Figure A.1.* Verbal memory (VM) latent growth modelling (LGM) estimated growth trajectory across 12 months for the control group ($n$ = 62).

**Latent Class Growth Analysis of the Verbal Memory Scores for the Control Group**

Table A.1, as previously noted, presents model fit indices from the 1- to 4-class growth models conducted incrementally. The model of best fit was selected from these statistical fit indices as well as for conceptual considerations. As shown in Table A.1, the results supported a 3-class model as this model had the lowest BIC and ABIC values (BIC = 884.532; ABIC = 846.777). Also, the LMR and adjusted LRT values for this model were significant, whereas those values for the 4-class model were not significant. The entropy value indicated that separation between the three classes was very good (0.869). It is noted that most participants were assigned to Classes 2 and 3 and that Class 1 contained only 3.56% of the control cohort. While the size for Class 1 was acceptable – at least 1% of the sample – the results have to be interpreted cautiously. Furthermore, the model demonstrated latent class probabilities of 1.00, 0.912 and 0.960 for Classes 1, 2 and 3, respectively, thereby adding further support for the 3-class model. There was also conceptual support for the 3-class model, based on the distinct trajectories of ARCD demonstrated in past studies (*e.g.*, Langbaum et al., 2009), which reported three distinct categories of global cognitive trajectories over a 15-year period in a cohort of elderly community-dwelling women.

**Descriptive Variables of Classes for the Control Group**

Table A.2 shows the mean (*M*, *SD*) scores for the three classes in the optimal LGCA model for baseline age, sex, number of years of education and estimated pre-morbid (WTAR) IQ scores. It also shows the results of the appropriate statistics (one-way ANOVA for comparison of age, years of education and pre-morbid IQ, and Pearson's $\chi^2$ for sex) and the results of planned comparisons between the classes on each these variables when the ANOVA revealed a statistical difference. As shown in Table A.2, the ANOVA results indicated significant class differences for age ($p$ = .004) and estimated premorbid IQ ($p$ = .02). There were no significant differences between the classes for years of education ($p$ = .453) nor sex ($p$ = .07). Planned comparisons of significant results indicated that for age, Class 1 was significantly older than Class 3, with a large difference size (Cohen's $d$ = 1.22). In addition, Class 1 had lower estimated IQ scores than Classes 2 and 3. Classes 2 and 3 did not differ from each other in IQ scores. Again the size of these differences between Class 1 and Classes 2 and 3 was large (Cohen's $d$ = –2.67 and –2.26, respectively). It is also worth noting that despite having lower estimated IQ, overall Class 1 had an average IQ score; the other two classes had high average scores.

Table A.2.

*Descriptive Variables of the Classes of the Verbal Memory (VM) 3-Class Model for the Control Group*

| | **Participant class** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Class 1 ($n$ = 2)** | | **Class 2 ($n$ = 20)** | | **Class 3 ($n$ = 40)** | | | | | | |
| **Predictor** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***F (df)*** | ***p*** | **Pairwise comparisons** | ***p*** | **Cohen's *d*** |
| Age (years) | 76.50 | 6.36 | 68.63 | 7.40 | 63.90 | 6.41 | 5.944 (2, 59) | .004** | 1 *vs.* 2 | .35 | 1.22 |
| | | | | | | | | | 1 *vs.* 3 | .04* | 1.96 |
| | | | | | | | | | 2 *vs.* 3 | .04* | 0.59 |
| Female (%) | 50.12 | | 63.23 | | 87.81 | | $\chi^2$ (2, $N$ = 62) = 8.293 [a] | .07 | | | Cramer's $V$ = 0.37 |
| Education (years) | 12.00 | 4.24 | 16.10 | 3.55 | 14.55 | 3.27 | 0.802 (2, 59) | .453 | | | |
| Estimated Premorbid IQ | 101.21 | 4.24 | 114.00 | 5.43 | 113.33 | 6.43 | 7.723 (2, 59) | .02* | 1 *vs.* 2 | .02* | –2.67 |
| | | | | | | | | | 1 *vs.* 3 | 1.00 | –2.26 |
| | | | | | | | | | 2 *vs.* 3 | .02* | 0.11 |

[a] Pearson's $\chi^2$ test. * $p$ < .05, ** $p$ < .01, two-tailed.

## Performance Trajectories of Classes for the Control Group

Table A.3 shows the parameters for the classes (intercepts and slopes). Class 1, the smallest class, had the lowest intercept, Class 2 had a higher intercept value than Class 1, and Class 3 had a higher intercept value than Class 2. Thus the initial VM performance of Class 1 was lower than Class 2 and Class 3, and Class 2 had lower initial VM performance than Class 3. One-way ANOVA (*e.g.*, Stulz et al., 2010) revealed a significant difference in intercept values between the classes ($F(2, 59) = 26.79$, $p < .001$). *Post-hoc* comparisons using Tukey HSD (honest significant difference) test indicated a significant difference between Class 1 and classes 2 and 3 ($p = .014$ and $p < .001$, respectively). There was also a significant difference between Class 2 and Class 3 ($p < .001$). Classes 1, 2 and 3 are referred to in the present study as Low, Moderate and High VM classes, respectively. In relation to the slope parameters (Figure A.2), the Low VM class showed a significant negative slope (Estimate = -.173, *SE* 0.063, $p = .006$). In contrast the slopes of the Moderate and High VM classes were not significant. The findings indicated that the VM performance of the Low VM class decreased over the 12-month follow-up period, compared with the moderate and high VM classes, which remained steady.

Table A.3.
*Growth Parameter Estimates for the Classes in the Verbal Memory (VM) 3-Class Model for the Control Group*

| **Variable** | **Estimate** | ***S.E.*** | ***p*** |
|---|---|---|---|
| 1: Low VM (*n* = 2) | | | |
| Intercept | 9.614 | 0.193 | <.001** |
| Slope | –0.173 | 0.063 | .006* |
| 2: Moderate VM (*n* = 19) | | | |
| Intercept | 11.612 | 0.353 | <.001** |
| Slope | 0.074 | 0.043 | .09 |
| 3: High VM (*n* = 41) | | | |
| Intercept | 13.882 | 0.175 | <.001** |
| Slope | 0.029 | 0.019 | .131 |

*Note:* 1, 2, 3 indicates model class assignment in model as per Table A.4. * $p < .01$, ** $p < .001$, two-tailed.

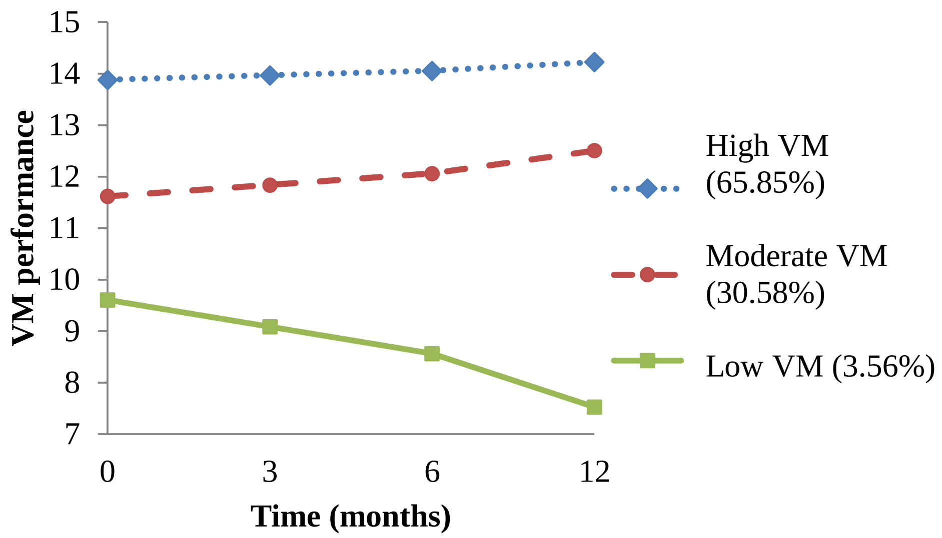


*Figure A.2.* Trajectories of the classes for verbal memory (VM) latent class growth analysis (LCGA) across 12 months for the control group (*n* = 62).

## Treatment Group

### Latent Growth Modelling of the Verbal Memory Scores for the Treatment Group

Table A.4 reveals the various fit indices of the LCGA models (and LGM) with variances held at zero to determine the optimal model of VM performance for the treatment group. Again, MPlus parameter default options were used. The single class model demonstrated the experimental group growth trajectory (model fitting indicators were BIC = 3089.019 and ABIC = 3069.999). The model revealed a significant positive slope for the treatment group ($n$ = 251, estimate = 0.087, $SE$ = 0.016, $p$ < .001), representing VM performance improvement across the 12-month period. Figure A.3 shows this improvement was demonstrated at the 6-month time-point, *i.e.*, not immediately following training. A LGM in which variances were allowed to vary (*e.g.*, Muthén et al., 2002) produced a negative slope variance value. As such, this model was rejected and subsequent models used fixed variances to demonstrate the treatment group VM trajectories.

Table A.4.
*Model Fit Indices from the Verbal Memory (VM) Growth Modelling for the Treatment Group*

| | | | LMR | Adjusted | | Class membership (%) | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Model** | **BIC** | **ABIC** | ***p*** | **LRT *p*** | **Entropy** | **C1** | **C2** | **C3** | **C4** |
| 1-class | 3089.019 | 3069.999 | | | | 100 | | | |
| 2-class | 2992.034 | 2963.503 | .0001* | .0002* | 0.625 | 38.54 | 61.46 | | |
| **3-class** | **2960.315** | **2922.274** | **.07** | **.07** | **0.718** | **5.63** | **52.25** | **42.12** | |
| 4-class | 2958.478 | 2910.926 | .14 | .15 | 0.720 | 11.13 | 40.60 | 46.97 | 1.23 |

*Note:* Bold indicates best fit. BIC = Bayesian information criterion; ABIC = sample-size adjusted Bayesian Information Criterion; LMR = Vuong-Lo-Mendell-Rubin likelihood ratio test; Adjusted LRT = Lo-Mendell-Rubin Adjusted likelihood ratio test. * $p$ < .001, two-tailed.

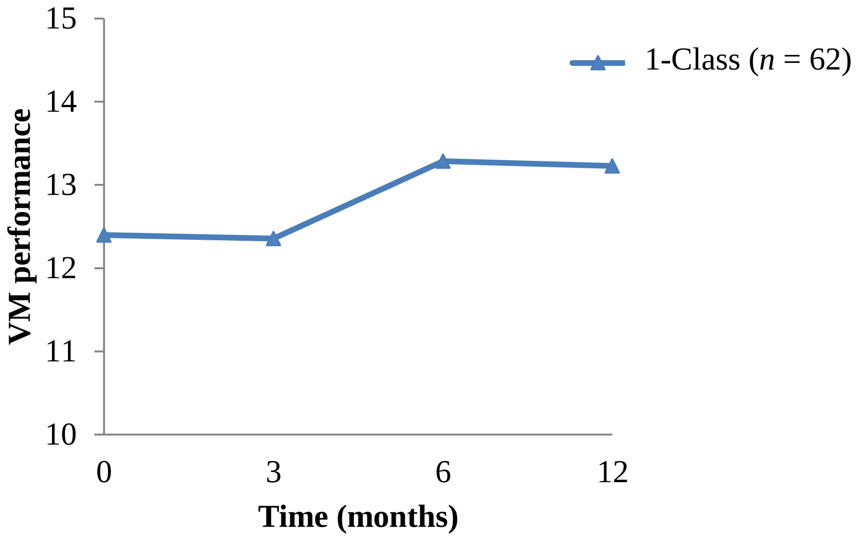


*Figure A.3.* Verbal memory (VM) latent growth modelling (LGM) estimated growth trajectory across 12 months for the treatment group ($n$ = 251).

**Latent Class Growth Analysis Model of the Verbal Memory Scores for the Treatment Group**

Table A.4 shows the model fit indices of the LCGA models executed to determine the optimal model of VM performance for the treatment group. Model assessment was performed iteratively from 1 to 2, 3 and 4 classes as it was for the control group. Again, the model of best fit was selected from these statistical fit indices. Conceptual considerations were also taken into account. Results supported a 3-class model. Whilst the 4-class model demonstrated the lowest BIC and AIC values (BIC = 2958.478; ABIC = 2910.926), the LMR and adjusted LRT values were not significant ($p$ = .15). Entropy of the 3-classes model was satisfactory (0.718). Again, Class 1 in the 3-class model was small (5.63% of sample), yet considered of adequate size. The average latent class probabilities calculated were 0.833, 0.893 and 0.823 for Classes 1, 2 and 3, respectively. There was also conceptual support for the 3-class model, based on the distinct trajectories of individuals undertaking memory training demonstrated in past studies that also demonstrated inter-individual variability in treatment response (*e.g.*, Willis et al., 2006).

**Descriptive Variables of Classes for the Treatment Group**

Table A.5 presents the mean (*SD*) scores for the three LCGA classes for baseline age, sex, number of years of education and pre-morbid (WTAR) IQ scores. It also shows the results of the appropriate statistics (one-way ANOVA for comparison of age, years of education and pre-morbid IQ, and Pearson's $\chi^2$ for sex) and the results of planned comparisons of the classes on each of these variables when ANOVA revealed statistical a difference. The results indicated significant class differences for age ($p = .001$), sex ($p < .001$) and years of education ($p = .01$). There was a trend towards a significant difference between the classes for estimated premorbid IQ ($p = .06$), although all classes can be considered to have above average IQ. Planned comparisons of significant results indicated that for age, Class 1 was significantly older than the Class 3 ($p = .03$), with a large difference (Cohen's $d = 0.78$). Class 1 also had the lowest percentage of females (41.67%). Of note, there was no significant difference in years of education between Class 1 and Classes 2 and 3. Class 2 was significantly older than Class 3, with a moderate difference ($p = .004$; Cohen's $d = 0.41$).

Table A.5.
*Descriptive Variables of the Classes from the Verbal Memory (VM) 3-Class Model for the Treatment Group*

| | Participant class | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Class 1 ($n$ = 12)** | | **Class 2 ($n$ = 110)** | | **Class 3 ($n$ = 129)** | | | | **Pairwise** | | |
| **Predictor** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***F (df)*** | ***p*** | **comparisons** | ***p*** | **Cohen's *d*** |
| Age (years) | 70.67 | 7.32 | 68.16 | 7.42 | 65.32 | 6.42 | 7.206 (2, 248) | .001** | 1 *vs.* 2 | .71 | 0.34 |
| | | | | | | | | | 1 *vs.* 3 | .03* | 0.78 |
| | | | | | | | | | 2 *vs.* 3 | .004** | 0.41 |
| Female (%) | 41.67 | | 72.72 | | 91.47 | | $\chi^2$ (2, $N$ = 251) = 24.287 [a] | <.001*** | | | Cramer's $V$ = 0.31 |
| Education (years) | 13.17 | 2.55 | 13.06 | 2.87 | 14.11 | 2.79 | 4.718 (2, 248) | .01* | 1 *vs.* 2 | .009** | 0.04 |
| | | | | | | | | | 1 *vs.* 3 | .73 | –0.35 |
| | | | | | | | | | 2 *vs.* 3 | 1.00 | –0.37 |
| Estimated Premorbid IQ | 111.02 | 5.55 | 110.45 | 7.26 | 112.48 | 5.89 | 2.864 (2, 247) | .06 | | | |

[a] Pearson's $\chi^2$ test. * $p$ < .05, ** $p$ < .01, *** $p$ < .001, two-tailed.

## Performance Trajectories of Classes for the Treatment Group

Table A.6 shows the growth parameters for the Classes (intercepts and slopes). Class 1 contained the lowest proportion of participants and had the lowest intercept; Class 2 had the highest intercept value and the greatest proportion of participants; and Class 3 had the second highest proportion of participants and intercept value. Thus the initial VM performance of Class 1 was lower than Class 2 and 3, and Class 3 had lower initial VM than Class 2. One-way ANOVA revealed that there was a significant difference in intercept values between all the classes ($F(2, 247) = 191.73$, $p < .001$). *Post-hoc* comparisons also indicated a significant difference between all three pairs of classes ($p < 0.001$). Classes 1, 2 and 3 are referred to as Low, VM and Moderate VM classes, respectively. When considering the slope parameters, all classes showed a significant positive slope ($p$ = .04, .007 and < .001 for Low, Moderate and High classes, respectively). Figure A.4 shows the trajectories of the slopes for each class.

The results indicated that the VM performances increased over the 12-month time interval for all treatment classes, with the Low VM class showing the greatest increase, followed by the Moderate class and then the High class. All classes demonstrated the highest estimated VM scores at the 12-month time point. As noted in the Method section, these models represented the standard model building process (Jung & Wickrama, 2008). The results were not used to make conclusions on VM improvements following training *vs.* control groups. Such conclusions were drawn from the joint analyses – *i.e.*, of the combined results for the control and treatment groups – described in the Chapter 7, Verbal Memory Performance.

Table A.6.
*Growth Parameter Estimates for Classes in the Verbal Memory (VM) 3-Class Model for the Treatment Group*

| **Variable** | **Estimate** | ***SE*** | ***p*** |
|---|---|---|---|
| 1: Low VM (*n* = 12) | | | |
| Intercept | 8.144 | 0.932 | <.001*** |
| Slope | 0.128 | 0.062 | .04* |
| 2: Moderate VM (*n* = 110) | | | |
| Intercept | 11.542 | 0.272 | <.001*** |
| Slope | 0.064 | 0.022 | .007** |
| 3: High VM (*n* = 129) | | | |
| Intercept | 13.554 | 0.143 | <.001*** |
| Slope | 0.075 | 0.016 | <.001*** |

*Note:* 1, 2, 3 indicates model class assignment in model as per Table A.5. * $p < .05$, ** $p < .01$, *** $p < .001$, two-tailed.

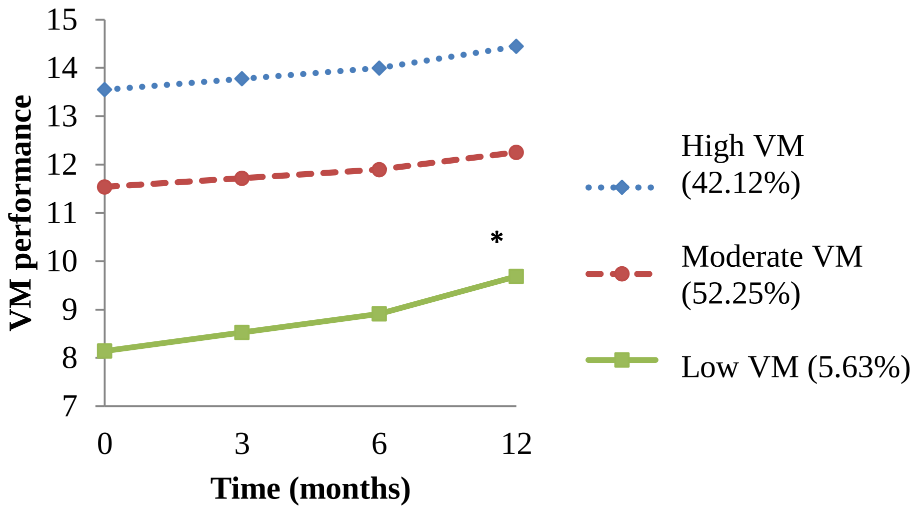


*Figure A.4.* Trajectories of the latent class growth analysis (LCGA) classes across 12 months for the treatment group (*n* = 251).

# APPENDIX 7

# Control and Treatment Groups' Model Results for Long-term Verbal Memory

## Control Group Long-term Verbal Memory Performance

### Latent Growth Modelling of the Long-term Verbal Memory Scores for the Control Group

Table A.7 presents model fit indices from the 1- to 4-class growth models conducted incrementally (the LGM and the LCGA) to identify the optimal LTVM performance model for the control group. As noted in the data analysis section, MPlus parameter default options were used.

As can be seen in the table, the single-class model for the control group in which variances were fixed resulted in a BIC = 1122.302 and ABIC = 1103.424. The LTVM trajectory is identified in Figure A.5. The largest point of change from baseline can be seen at 12 months. There was a significant positive growth in LTVM performance ($n$ = 62, Estimate = 0.08, $SE$ = 0.04, $p$ = .04). A LGM in which variances were allowed to vary (*e.g.*, Muthén & Muthén, 2000) produced a model indicating significant positive growth ($n$ = 62, Estimate = 0.08, $SE$ = 0.03, $p$ = .001). It should be noted, however, that variances of the slope parameter were not significant ($p$ = .43). As such, subsequent control models were conducted with variances fixed.

Table A.7.
*Model Fit Indices from the Long-term Verbal Memory (LTVM) Latent Growth Modelling (LGM) for the Control Group*

| | | | LMR | Adjusted | | Class membership (%) | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Model** | **BIC** | **ABIC** | ***p*** | **LRT *p*** | **Entropy** | **C1** | **C2** | **C3** | **C4** |
| 1-class | 1122.302 | 1103.424 | | | | 100 | | | |
| 2-class | 1044.429 | 1016.112 | . 09 | .10 | .86 | 37.10 | 62.90 | | |
| **3-class** | **1023.932** | **986.177** | **.11** | **.12** | **.89** | **14.51** | **44.01** | **41.48** | |
| 4-class | 1036.314 | 989.119 | .82 | .82 | .91 | 14.51 | 41.48 | 44.01 | 0.004 |

*Note:* Bold indicates best fit. BIC = Bayesian information criterion; ABIC = sample-size adjusted Bayesian Information Criterion; LMR = Vuong-Lo-Mendell-Rubin likelihood ratio test; Adjusted LRT = Lo-Mendell-Rubin Adjusted likelihood ratio test.

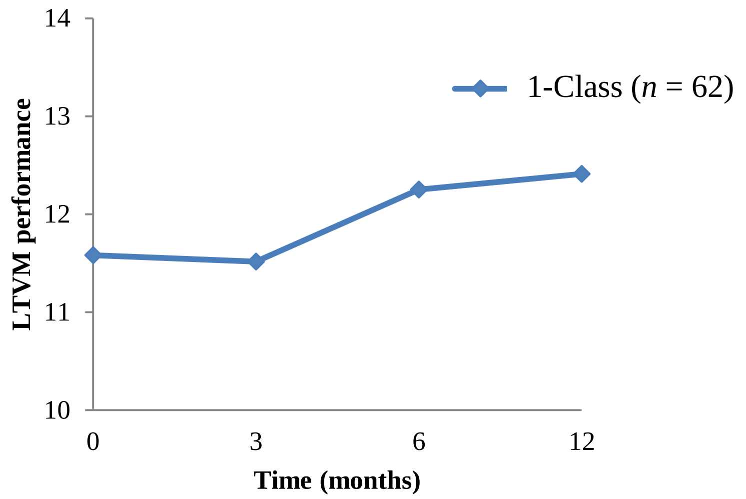


*Figure A.5.* Long-term verbal memory (LTVM) latent growth modelling (LGM) estimated growth trajectory across 12 months for the control group (*n* = 62).

## Latent Class Growth Analysis of the Long-term Verbal Memory Scores for the Control Group

As previously noted, Table A.7 presents model fit indices from the 1- to 4-class growth models. The model of best fit was selected from these statistical fit indices as well as conceptual considerations. As shown in Table A.7, the results supported a 3-class model. Table A.7 demonstrates that the 3-class model had the lowest BIC and ABIC values (BIC = 1023.932; ABIC = 986.177). Whilst the LMR and adjusted LRT values for this model were not significant (*p* = .11 and .12, respectively), the entropy value indicated that separation between the three classes was very good (0.89) and the sizes of the classes were acceptable (14.51%, 44.01% and 41.48%, respectively). Furthermore, the model demonstrated latent class probabilities of 0.965, 0.981 and 0.923 for Classes 1, 2 and 3, respectively, thereby adding further support for the 3-class model. There was also conceptual support for the 3-class model, based on reports of 3 distinct trajectories of cognitive performances (*e.g.*, Langbaum et al., 2009; as outlined in Chapter 4).

**Descriptive Variables of Classes for the Control Group**

Table A.8 shows the mean (*SD*) scores for the three LGCA classes for baseline age, sex, number of years of education and estimated pre-morbid (WTAR) IQ scores. It also shows one-way ANOVA results for comparison of age, years of education and pre-morbid IQ, and Pearson's $\chi^2$ for sex. As shown in the table, the results of the ANOVA indicated no significant class differences for age (*p* = .47), years of education (*p* = .51) nor estimated premorbid IQ (*p* = .19). Individuals of all classes could be considered to be, on average, young-older adults (APA, 2009) and tertiary educated. All classes also showed a High Average estimated pre-morbid IQ compared with norms (Sattler & Dumont, 2004). There was, however, a significant difference in the number of females (*p* = .01) between the classes, which can be considered of a medium-large size (Cramer's *V* = 0.39; Gravetter & Wallnau, 2004).

Table A.8.

*Descriptive Variables of the Classes of the Long-term Verbal Memory (LTVM) 3-Class Model for the Control Group*

| | Participant class | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Class 1 ($n$ = 9) | | Class 2 ($n$ = 27) | | Class 3 ($n$ = 26) | | | | |
| | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***F (df)*** | ***p*** | **Cramer's *V*** |
| Age (years) | 68.33 | 7.18 | 65.74 | 7.98 | 64.88 | 6.34 | 0.76 (2,59) | 0.47 | |
| Female (%) | 44.40 | | 77.80 | | 92.30 | | $\chi^2$ (2, $N$ = 62) = 9.29 | .001[a] | 0.39 |
| Education (years) | 14.56 | 3.47 | 14.34 | 3.39 | 15.44 | 3.66 | 0.68 (2, 59) | 0.51 | |
| Estimated Premorbid IQ | 112.33 | 7.16 | 111.74 | 7.34 | 114.88 | 4.693 | 1.72 (2, 59) | 0.19 | |

[a] Pearson's $\chi^2$ test, * $p$ = .001, two-tailed.

**Performance Trajectories of Classes for the Control Group**

Table A.9 shows the parameters (intercepts and slopes) for the three classes. Class 1, the smallest class, had the lowest intercept; Class 2, the largest class had a higher intercept value than Class 1, although Class 3, the second largest class had the highest intercept value of all three classes. Thus the initial long-term memory performance of Class 1 was lower than Classes 2 and 3, and Class 2 had lower initial LTVM performance than Class 3. One-way ANOVA revealed that there was a significant difference in intercept values between the classes ($F(2, 59) = 99.52$, $p < .001$). *Post-hoc* comparisons using Tukey HSD indicated that there was a significant difference between all classes ($p < .001$). Classes 1, 2 and 3 are therefore referred to as Low, Moderate and High LTVM classes, respectively. Table A.9 also shows the slope parameters. The Low LTVM class showed a significant negative slope (Estimate = 0.246, $SE = 0.111$, $p = .03$). In contrast the slopes of the Moderate and High LTVM classes were not significant ($p = .07$ .40, respectively). Figure A.6 shows the trajectories of the slopes for each class. Overall, the results indicated that the LTVM performance of the Low LTVM class increased over the 12-month follow-up period, whereas the LTVM performance of the Moderate and High LTVM classes remained steady over this interval.

Table A.9.

*Growth Parameter Estimates for the Classes in the Long-term Verbal Memory (LTVM) 3-Class Model for the Control Group*

| **Variable** | **Estimate** | ***SE*** | ***p*** |
|---|---|---|---|
| 1: Low LTVM (*n* = 9) | | | |
| Intercept | 6.872 | 0.356 | <.001** |
| Slope | 0.246 | 0.111 | .03* |
| 2: Moderate LTVM (*n* = 27) | | | |
| Intercept | 10.631 | 0.443 | <.001** |
| Slope | 0.095 | 0.052 | .07 |
| 3: High LTVM (*n* = 26) | | | |
| Intercept | 13.774 | 0.197 | <.001** |
| Slope | 0.022 | 0.022 | .40 |

*Note:* 1, 2, 3 indicates model class assignment in model as per Table A.8. * $p < .05$, ** $p < .001$, two-tailed.

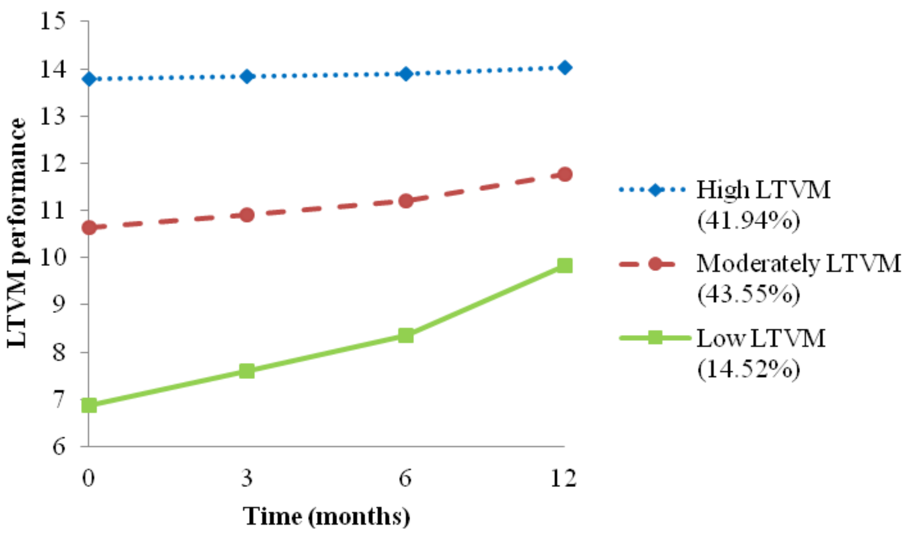


*Figure A.6.* Trajectories of the classes for long-term verbal memory (LTVM) latent growth class analysis (LCGA) across 12 months for the control group (*n* = 62).

## Treatment Group

### Latent Growth Modelling of the Long-term Verbal Memory Scores for the Treatment Group

Table A.10 reveals the various fit indices of the growth models with variances held at zero, conducted to determine the optimal model of LTVM performance for the treatment group. As noted in the data analysis section of the Method chapter, MPlus parameter default options were used. The single class model demonstrated the experimental group growth trajectory (the model fitting indicators were BIC = 3645.38 and ABIC = 3626.359; see Table A.10). The model revealed a significant positive slope for the treatment group ($n$ = 251, Estimate = 0.16, $SE$ = 0.02, $p$ ≤.001). Figure A.7 shows the greatest level of improvement from baseline was demonstrated at the 6-month time-point (*i.e.*, not immediately following training). A LGM in which variances were allowed to vary (*e.g.*, Muthén & Muthén, 2000) produced a negative slope variance value. As such this model was inadmissible and therefore was rejected. As previously highlighted, subsequent models used fixed variances to demonstrate the treatment group LTVM trajectories.

Table A.10.
*Model Fit Indices from Long-term Verbal Memory (LTVM) Growth Modelling for the Treatment Group*

| | | | **LMR** | **Adjusted LRT** | | **Class membership (%)** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **BIC** | **ABIC** | ***p*** | ***p*** | **Entropy** | **C1** | **C2** | **C3** | **C4** | **C5** |
| 1-class | 3645.382 | 3626.359 | | | | 100 | | | | |
| 2-class | 3552.966 | 3524.435 | .01** | .02* | 0.627 | 37.45 | 62.55 | | | |
| **3-class** | **3525.746** | **3487.704** | **.06** | **.07** | **0.665** | **47.01** | **45.02** | **7.97** | | |
| 4-class | 3524.155 | 3476.603 | .02* | .02* | 0.719 | 5.18 | 49.8 | 43.43 | 1.6 | |
| 5-class | 3527.426 | 3470.363 | .27 | .29 | 0.65 | 0.06 | 0.18 | 0.32 | 0.02 | 0.41 |

*Note:* Bold indicates best fit. BIC = Bayesian information criterion; ABIC = sample-size adjusted Bayesian Information Criterion; LMR = Vuong-Lo-Mendell-Rubin likelihood ratio test; Adjusted LRT = Lo-Mendell-Rubin Adjusted likelihood ratio test. * $p$ < .05, ** $p$ = .01, two-tailed.

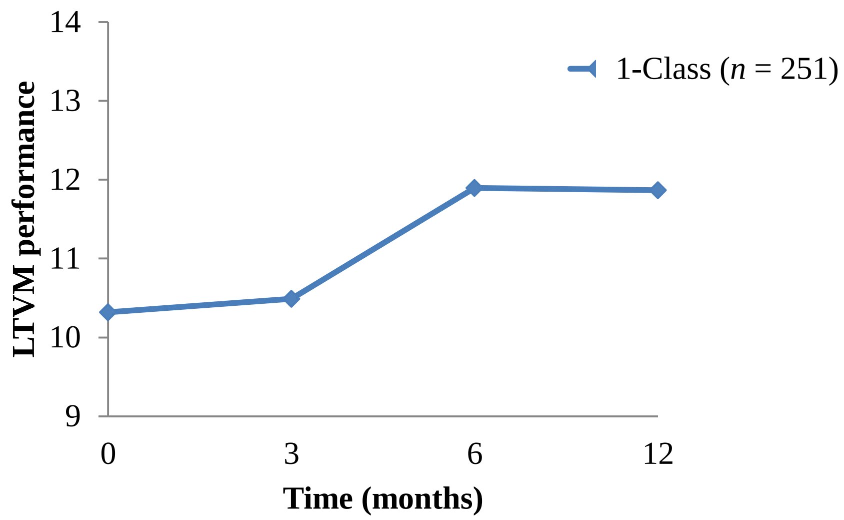


*Figure A.7.* Long-term verbal memory (LTVM) latent growth modelling (LGM) estimated mean growth trajectory across 12 months for the control group ($n$ = 251).

**Latent Class Growth Analysis Model of the Long-term Verbal Memory Scores for the Treatment Group**

Table A.10 also shows the model fit indices of the LCGA models executed to determine the optimal model of LTVM performance for the treatment group. As was conducted for the control group, model assessment was performed iteratively from 1 to 2, 3 and 4 classes. Again, the model of best fit was selected from these statistical fit indices. Conceptual considerations were also taken into account. Whilst the 4-class model demonstrated the lowest BIC and ABIC values model (BIC = 3524.155; ABIC = 3476.603), and produced significant LMR and adjusted LRT values ($p$ = .02), supporting a 4-class model, the model was discounted in favour of the 3-class model for a number of reasons. Class 4 in the 4-class model is small (1.6% of sample) and, whilst it was deemed statistically acceptable, the classes were not qualitatively distinct (*e.g.*, Uher et al. 2010). Specifically, one-way ANOVA ($F(2, 245) = 98.87$, $p < .001$) with Bonferroni *post-hoc* comparisons revealed baseline performance between Classes 1 and 4, and Classes 2 and 4 were not significant ($p$ = .12 and = .09, respectively). Considering the 3-class model, one-way ANOVA with Bonferroni *post-hoc* comparisons ($F(2, 246) = 164.44$, $p < .001$) revealed significant differences between all classes in the 3-class model (all comparisons $p < .001$). The 3-class model also had significant likelihood ratio tests (LMR $p = .06$ and adjusted LRT $p = .07$), the entropy was satisfactory (0.665), and the classes were considered of adequate size – 47.01%, 45.02% and 7.97%, respectively. The average latent class probabilities calculated were 0.85, 0.84 and 0.85, for Classes 1, 2 and 3, respectively. Furthermore, there was conceptual support for the 3-class model. Firstly it matched the number of classes in the control group previously identified in the present study, as is considered class enumeration confirmation (*cf.* Muthén et al., 2002). Secondly, past studies have demonstrated three distinct trajectories of individuals undertaking memory training (*e.g.*, Willis et al., 2006).

**Descriptive Variables of Classes for the Treatment Group**

Table A.11 presents the mean (*SD*) scores for the three LCGA classes for baseline age, sex, number of years of education and pre-morbid (WTAR) IQ scores. It also shows the results of the appropriate statistical analyses (one-way ANOVA for comparison of age, years of education and pre-morbid IQ, and Pearson's $\chi^2$ for sex) and a planned comparison of the classes on each of these variables when ANOVA revealed a statistical difference. The results shown in the table indicated significant differences between classes for age ($p = .001$) and sex ($p < .001$). The different proportion of females in each class was considered to be of medium size (Cramer's $V = 0.32$). There were no significant differences between the classes for education or estimated premorbid IQ ($p = .53$ and .32, respectively). Individuals of all classes could be considered as having had, on average, some tertiary education and a High Average estimated IQ. Planned comparisons of significant results indicated that for age, Class 1 was significantly older than Class 3 ($p$ = .02), with a large difference (Cohen's $d$ = 0.87). Class 1 also had the lowest percentage of females (45.00%).

Table A.11.

*Descriptive Variables of the Classes from the Long-term Verbal Memory (LTVM) 3-Class Model for the Treatment Group*

| | Participant class | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Class 1 (*n* = 118)** | | **Class 2 (*n* = 113)** | | **Class 3 (*n* = 12)** | | | | | | |
| **Predictor** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***F (df)*** | ***p*** | **Pairwise comparisons** | ***p*** | **Cohen's *d*** |
| Age (years) | 70.95 | 6.91 | 65.2 | 6.33 | 67.78 | 7.45 | 7.55 (2, 248) | 0.001 | 1 *vs.* 2 | .02* | 0.869 |
| | | | | | | | | | 1 *vs.* 3 | .16 | 0.442 |
| | | | | | | | | | 2 *vs.* 3 | .002** | –0.374 |
| Female (%) | 77.71 | | 91.20 | | 45.00 | | $\chi^2$ (2, *N* = 251) = 25.43 | <.001[a] | | <.001[a] | Cramer's *V* = 0.32 |
| Education (years) | 13.23 | 3.41 | 13.76 | 2.84 | 13.38 | 2.76 | 0.64 (2, 248) | 0.53 | | | |
| Estimated Premorbid IQ | 111.35 | 5.2 | 112.21 | 6.1 | 110.9 | 7.21 | 1.16 (2, 248) | 0.32 | | | |

[a] Pearson's $\chi^2$ test. * $p < .05$, ** $p < .01$, ***.$p \leq .001$, two-tailed.

**Performance Trajectories of Classes for the Treatment Group**

Table A.12 shows the growth parameters for the classes (intercepts and slope). Class 1 contained the highest proportion of participants and had an intercept that was between Classes 2 and 3. Class 2 had the highest intercept value and had the second greatest proportion of participants. Class 3 had the lowest proportion of participants and the lowest intercept value. Thus the order of the classes from lowest to highest initial LTVM performance was Class 3, 1 then 2. As previously noted, there was a significant difference between the intercept values for the three classes (all comparisons $p < .001$). Classes 1, 2 and 3 are referred to as Moderate, High and Low LTVM classes, respectively. When considering the slope parameters, the Moderate and Low LTVM classes showed a significant positive slope ($p \leq .001$). The High LTVM group did not show significant growth ($p < .48$). Figure A.8 shows the trajectories of the slopes for each class.

The results indicate that the LTVM performance trajectories increased over the 12-month time interval for the Low and Moderate LTVM classes, with the Low LTVM class showing the largest slope estimate. Individuals in the High LTVM class did not show improvement in performance. All classes demonstrated the highest estimated LTVM scores at the 12-month time point. As noted in the Method section, these models represented the standard model building process (Jung & Wickrama, 2008). The results were not used to make conclusions about LTVM improvements following training versus control groups. Such conclusions were drawn from the joint analyses (*i.e.*, of the combined results for the control and treatment groups), described in Chapter 8.

Table A.12.
*Growth Parameter Estimates for the Classes in the Long-term Verbal Memory (LTVM) 3-Class Model for the Treatment Group*

| **Variable** | **Estimate** | ***SE*** | ***p*** |
|---|---|---|---|
| 3: Low LTVM (*n* = 20) | | | |
| Intercept | 5.791 | 1.056 | <.001* |
| Slope | 0.173 | 0.024 | <.001* |
| 1: Moderate LTVM (*n* = 118) | | | |
| Intercept | 9.654 | 0.301 | <.001* |
| Slope | 0.092 | 0.026 | <.001* |
| 2: High LTVM (*n* = 113) | | | |
| Intercept | 12.072 | 0.226 | <.001* |
| Slope | 0.087 | 0.123 | .48 |

*Note:* 1, 2, 3 indicates model class assignment in model as per Table 24. * $p < .001$, two-tailed.

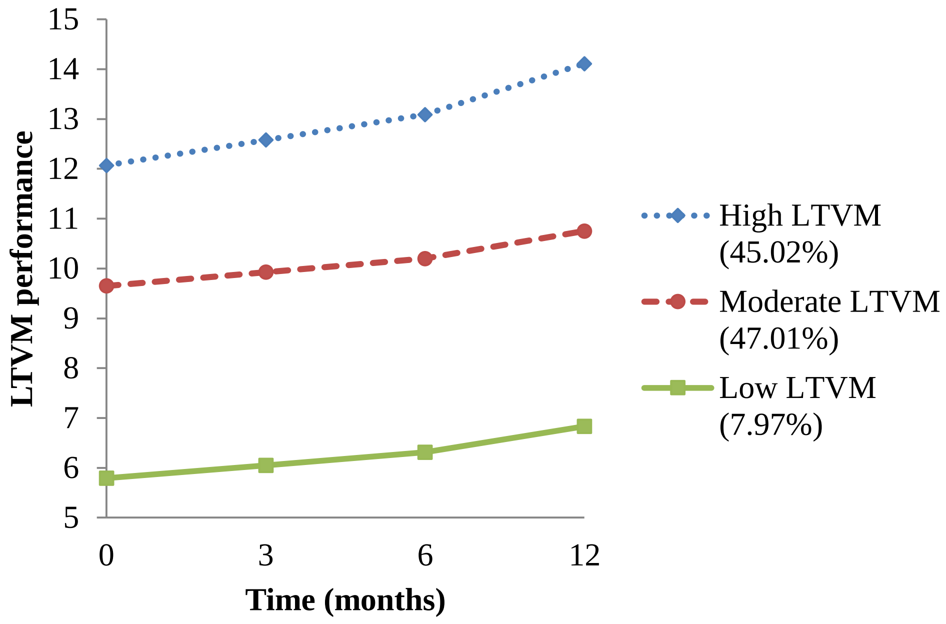


*Figure A.8.* Trajectories of the classes for the long-term verbal memory (LTVM) latent class growth analysis (LCGA) across 12 months for the treatment group (*n* = 251).

# APPENDIX 8

# Control and Treatment Groups' Model Results for Executive Function

## Control Group Executive Function Performance

### Latent Growth Modelling of the Executive Function Scores for the Control Group

Table A.13 presents model fit indices from the 1- to 4-class growth models conducted incrementally (the LGM and the LCGA) to identify the optimal EF performance model for the control group. As noted in the data analysis section, MPlus parameter default options were used. As can be seen in the table, the single-class model for the control group in which variances were fixed resulted in a BIC = 1960.113 and ABIC = 1941.235. The EF trajectory is identified in Figure A.9. The largest point of change (*i.e.*, the lowest error score) can be seen at 12 months. There was a significant negative growth in EF performance ($n$ = 62, Estimate = –0.98, $SE$ = 0.24, $p$ < .001). Thus, there was a significant improvement in EF performance (*i.e.*, fewest errors were made) across the 12-month interval.

A LGM in which variances were allowed to vary (*e.g.*, Muthén & Muthén, 2000) produced a model (BIC = 1873.283 and ABIC = 1844.966) indicating an admissible model that had significant negative growth ($n$ = 62, Estimate = –0.800, $SE$ = 0.151, $p$ < .001). It should be noted, however, that variances of the slope parameter were not significant ($p$ = 0.22). As such subsequent control models were conducted with variances fixed.

Table A.13.
*Model Fit Indices from the Executive Function (EF) Growth Modelling for the Control Group*

| | | | LMR | Adjusted | | Class membership (%) | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Model** | **BIC** | **ABIC** | ***p*** | **LRT *p*** | **Entropy** | **C1** | **C2** | **C3** | **C4** |
| 1-class | 1960.119 | 1941.245 | | | | 100 | | | |
| 2-class | 1911.257 | 1882.942 | 0.23 | 0.24 | 0.835 | 75.81 | 24.19 | | |
| **3-class** | **1887.633** | **1849.871** | **0.02*** | **0.03*** | **0.892** | **67.74** | **19.36** | **12.9** | |
| 4-class | 1890.414 | 1843.211 | 0.27 | 0.28 | 0.788 | 9.68 | 27.42 | 16.13 | 46.77 |

*Note:* Bold indicates best fit. BIC = Bayesian information criterion; ABIC = sample-size adjusted Bayesian Information Criterion; LMR = Vuong-Lo-Mendell-Rubin likelihood ratio test; Adjusted LRT = Lo-Mendell-Rubin Adjusted likelihood ratio test. * $p < .05$, two-tailed.

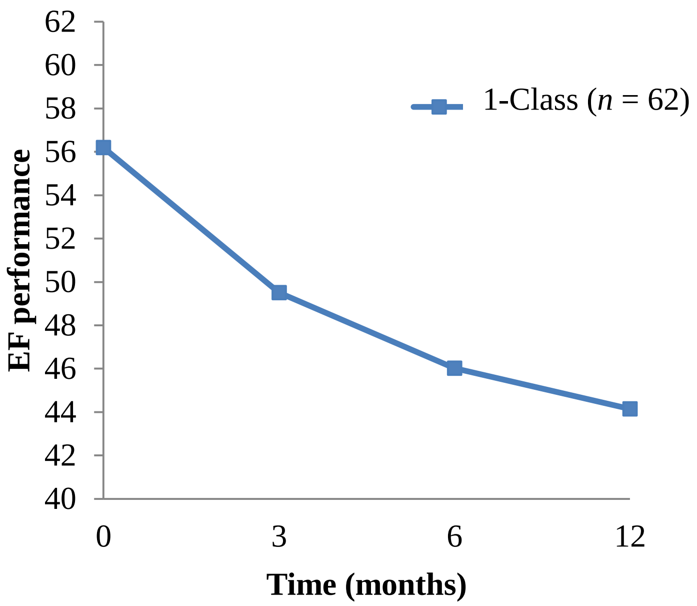


*Figure A.9.* Executive Function (EF) latent growth modelling (LGM) estimated growth trajectory across 12 months for the control group ($n = 62$).

## Latent Class Growth Analysis of the Executive Function Scores for the Control Group

As previously noted, Table A. 13 presents model fit indices from the 1- to 4-class growth models. The model of best fit was selected from these statistical fit indices as well as conceptual considerations. The results supported a 3-class model, as it had the lowest BIC and ABIC values (BIC = 1887.63; ABIC = 1849.87). The 3-class model also had significant LMR and adjusted LRT values ($p$ = .02 and .03, respectively), which were not significant for the 4-class model ($p$ = .27 and .28, respectively). The entropy value for the 3-class model indicated that separation between the three classes was very good (0.89) and the sizes of the classes were acceptable (67.74%, 19.36% and 12.9%, respectively). Furthermore, the model demonstrated latent class probabilities of 0.969, 0.912 and 0.922 for Classes 1, 2 and 3, respectively, thereby adding further support for the 3-class model. There was also conceptual support model, based on reports of three distinct trajectories of global cognitive performances (*e.g.*, Langbaum et al., 2009).

**Descriptive Variables of Classes for the Control Group**

Table A.14 shows the mean (*SD*) scores for the three LGCA classes for baseline age, sex, number of years of education and estimated pre-morbid (WTAR) IQ scores. It also shows one-way ANOVA results for comparison of age, years of education and pre-morbid IQ, and Pearson's $\chi^2$ for sex. As shown in the table, the results of the ANOVA indicated no significant class differences for age ($p$ = .99) or years of education ($p$ = .42). Individuals of all classes could be considered to be, on average, young-older adults (APA, 2009) and tertiary educated. There was a trend towards a significant difference between groups for estimated premorbid IQ ($p$ = .06). Individuals in all classes had a High Average estimated pre-morbid IQ compared with comparative norms when rounded to the nearest whole number. There was also no significant difference in the number of females ($p$ = .11) between the classes.

Table A.14.
*Descriptive Variables of the Classes of the Executive Function (EF) 2-Class Model for the Control Group*

| | Participant class | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Class 1 (*n* = 42)** | | **Class 2 (*n* = 12)** | | **Class 3 (*n* = 8)** | | | |
| | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***F (df)*** | ***p*** |
| Age (years) | 65.67 | 6.57 | 65.67 | 9.88 | 64.88 | 6.62 | 0.01 (2, 59) | 0.99 |
| Female (%) | 76.20 | | 100.00 | | 62.50 | | $\chi^2$ (2, *N* = 62) = 4.71 | 0.11 [a] |
| Education (years) | 14.70 | 3.65 | 14.33 | 3.45 | 16.32 | 2.61 | 0.87 (2, 59) | 0.42 |
| Estimated Premorbid IQ | 113.48 | 6.65 | 109.75 | 6.25 | 116.5 | 1.85 | 3.03 (2, 59) | 0.06 |

[a] Pearson's $\chi^2$ test, two-tailed.

**Performance Trajectories of Classes for the Control Group**

Table A.15 shows the parameters (intercepts and slopes) for the three classes. Class 1, the largest class, had an intercept value between Class 1 and 3. Class 2, the second largest class, had the highest intercept value, and Class 3 was the smallest class with the lowest intercept value. Thus the initial EF performance of Class 2 was the poorest of the three groups (*i.e.*, it had the highest number of errors). One-way ANOVA revealed that there was a significant difference in intercept values between the classes ($F(2, 59) = 44.46$, $p < .001$). *Post-hoc* comparisons using Tukey HSD indicated that there was a significant difference between all classes ($p < .001$). Classes 1, 2 and 3 are therefore referred to as Moderate, Low and High EF classes, respectively. Table 41 also shows the slope parameters. The Low EF class showed a significant negative slope (Estimate = –1.08, $SE = 0.46$, $p = .02$), as did the Moderate EF class (Estimate = –0.94, $SE = –0.17$, $p < .001$). In contrast, the slope of the High EF class was not significant ($p = .07$). A significant result may have been present with a larger number of participants in this class. Figure A.10 shows the trajectories of the slopes for each class. Overall, the results indicate that that the EF performance of the Low and Moderate EF classes increased over the 12-month follow-up period, whereas the EF performance of the High EF class remained steady over this time interval.

Table A.15.
*Growth Parameter Estimates for the Classes in the Executive Function (EF) 2-Class Model for the Control Group*

| **Variable** | **Estimate** | ***SE*** | ***p*** |
|---|---|---|---|
| 2: Low (*n* = 12) | | | |
| Intercept | 76.493 | 4.135 | <.001** |
| Slope | –1.077 | 0.46 | .02* |
| 1: Moderate (*n* = 42) | | | |
| Intercept | 52.166 | 2.047 | <.001** |
| Slope | –0.944 | –0.169 | <.001** |
| 3: High (*n* = 8) | | | |
| Intercept | 28.316 | 3.362 | <.001** |
| Slope | –0.302 | 0.165 | .07 |

*Note:* 1, 2, 3 indicates model class assignment in model as per Table A.13 and Table A.14.
* *p* < .05, ** *p* < .001, two-tailed.

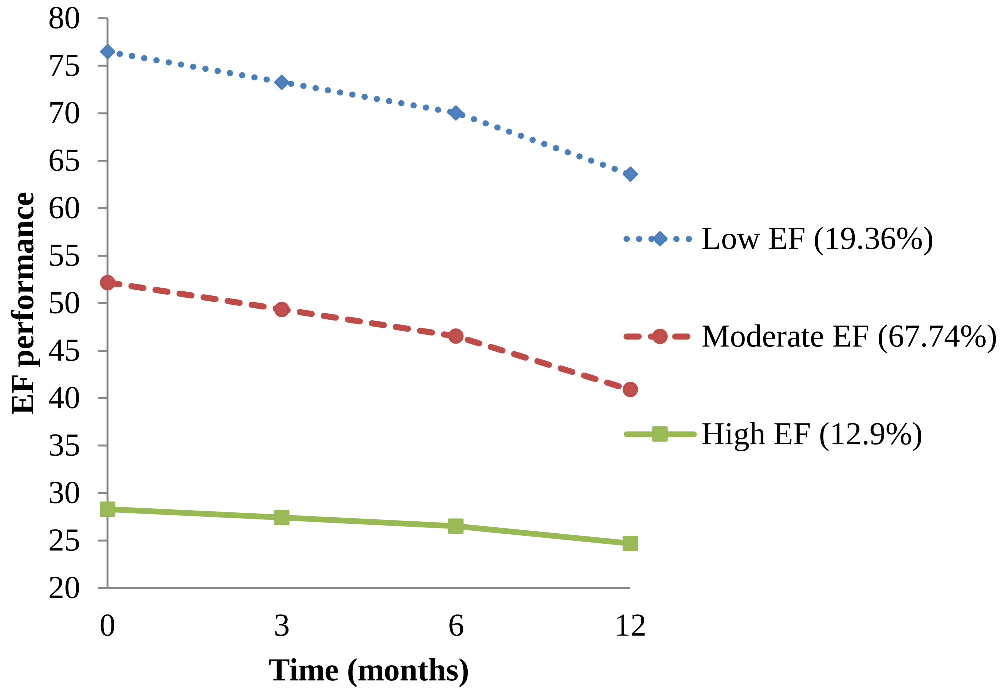


*Figure A.10.* Trajectories of the classes for executive function (EF) latent class growth analysis (LCGA) across 12 months for the control group (*n* = 62).

## Treatment Group

### Latent Growth Modelling of the Executive Function Scores for the Treatment Group

Table A.16 reveals the various fit indices of the LCGA models with variances held at zero and MPlus parameter default options used to determine the optimal model of EF performance for the treatment group. As noted in the data analysis section, MPlus parameter default options were used. The single class model (BIC = 6303.417 and ABIC = 6284.399) revealed a significant negative slope for the treatment group (*n* = 235, Estimate = –0.873, *SE* = 0.218, *p* < .001). This represented EF performance improvement across the 12-month period similar to the control groups' LGM previously identified (*n* = 62, Estimate = –0.98, *SE* = 0.24, *p* < .001). Figure A.11 shows this improvement was greatest in comparison to baseline at the 12-month time-point (*i.e.*, not immediately following training). A LGM in which variances were allowed to vary (*e.g.*, Muthén et al., 2002; BIC = 5976.216 and ABIC = 5947.690) also showed significant improvement in EF performance ($n$ = 62, Estimate = –0.927, $SE$ = 0.14, $p$ < .001). It should be noted, however, that variances of the slope parameter were not significant ($p$ = 0.07). As such, subsequent treatment models were conducted with variances fixed.

Table A.16.
*Model Fit Indices from Executive Function (EF) Growth Modelling for the Treatment Group*

| Model | BIC | ABIC | LMR *p* | Adjusted LRT *p* | Entropy | Class membership (%) C1 | C2 | C3 |
|---|---|---|---|---|---|---|---|---|
| 1-class | 6303.417 | 6284.399 | | | | 100 | | |
| **2-class** | **5995.258** | **5966.732** | **.004*** | **.006*** | **0.941** | **82.98** | **17.02** | |
| 3-class | 5943.766 | 5905.731 | .44 | .45 | 0.891 | 78.72 | 14.04 | 7.23 |

*Note:* Bold indicates best fit. BIC = Bayesian information criterion; ABIC = sample-size adjusted Bayesian Information Criterion; LMR = Vuong-Lo-Mendell-Rubin likelihood ratio test; Adjusted LRT = Lo-Mendell-Rubin Adjusted likelihood ratio test. * $p$ < .01, two-tailed.

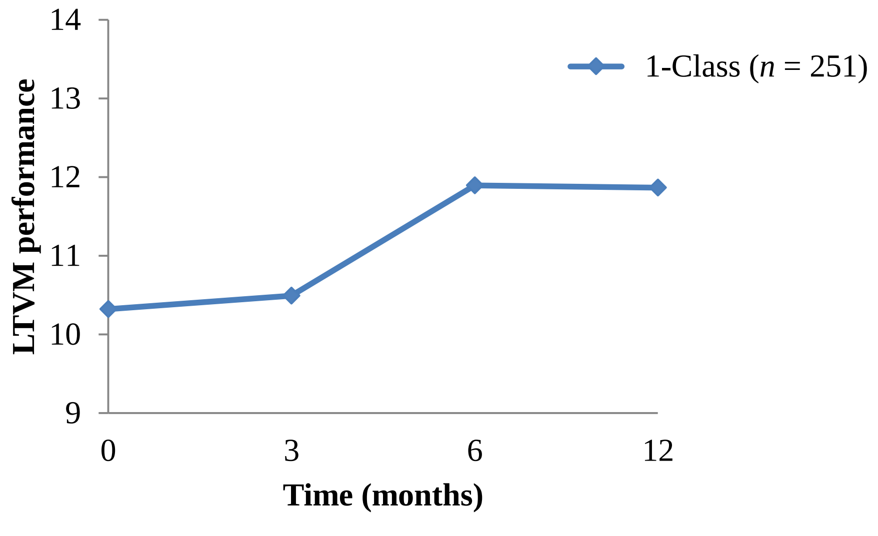


*Figure A.11*. Executive function (EF) latent growth modelling (LGM) estimated mean growth trajectory across 12 months for the treatment group ($n$ = 235).

**Latent Class Growth Analysis Model of Executive Function Scores for the Treatment Group**

Table A.16 also shows the model fit indices of the LCGA models executed to determine the optimal model of EF performance for the treatment group. Model assessment was performed iteratively from 1 to 3 classes. Again, the model of best fit was selected from these statistical fit indices as well as from conceptual considerations. Whilst the 3-class model demonstrated the lowest BIC and ABIC values model (BIC = 5943.766; ABIC = 5905.731), it did not produce significant LMR and adjusted LRT values ($p$ = .44 and .45, respectively). The 3-class model was therefore discounted in favour of the 2-class model (BIC = 5995.258; ABIC = 5966.732), which had significant LMR and adjusted LRT values ($p$ = .004 and .006, respectively). An independent-samples $t$-test revealed that there was a significant difference between Class 1 ($M$ = 50.38, $SD$ = 14.48) and Class 2 ($M$ = 106.75, $SD$ = 22.53) in the 2-class model $t(223) = -19.35$, $p < .001$) at the initial time-point. The 2-class model also had excellent entropy (0.94), and the classes were considered of adequate size (82.98% and 17.02%, respectively). The average latent class probabilities calculated were 0.993 and 0.947 for Classes 1 and 2, respectively. Conceptually, a 2-class model did not fit the control group, or past studies that demonstrated three distinct trajectories in global measures of cognitive functioning (*e.g.*, Langbaum et al., 2009). However, given that there have been no other studies to date specifically investigating EF trajecotories following training results, theoretically a 2-class model for EF was possible. Results of this model were also easily interpretable with the presence of qualitiatively distinct groups based on initial performance levels (*e.g.*, Uher et al. 2010). Together with strong statistical support, this model was therefore considered the optimal model to highlight the heterogeneity of EF trajectories of participants following training.

### Descriptive Variables of Classes for the Treatment Group

Table A.17 presents the mean (*SD*) scores for the two LCGA classes for baseline age, sex, number of years of education and pre-morbid (WTAR) IQ scores. It also shows the results of the appropriate statistics: independent-samples *t*-tests for comparison of age, years of education and pre-morbid IQ, and a $\chi^2$ test for independence for sex with Yate's Continuity Correction. As shown in the table, the results indicated significant differences between the two classes for age ($p$ = .001) and estimated premorbid IQ ($p$ = .002). These differences can be considered to be medium and small (Cohen's $d$ = – 0.66 and 0.33, respectively). Despite the difference in age between those in the two classes, they are both considered younger-older adults (APA, 1998). Individuals in the classes were High Average and Average estimated IQ, respectively. There were no differences in the proportion of females, nor differences for education ($p$ = .52 and .14, respectively).

Table A.17.
*Descriptive Variables of the Classes from the Executive Function (EF) 2-Class Model for the Treatment Group*

| | **Participant class** | | | | | |
|---|---|---|---|---|---|---|
| | **Class 1 ($n$ = 195)** | | **Class 2 ($n$ = 40)** | | | |
| | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***t (df)*** | ***p.*** |
| Age (years) | 65.97 | 6.79 | 70.57 | 7.13 | -3.86 (233) | <.001** |
| Female (%) | 81.44 | | 75.21 | | $\chi^2$ (1, $N$ = 235) = .42 | 0.52 [a] |
| Education (years) | 13.75 | 2.79 | 13.03 | 3.00 | 1.48 (233) | 0.14 |
| Estimated Premorbid IQ | 112.24 | 5.681 | 104.05 | 34.18 | 3.17 (233) | 0.002* |

[a] $\chi^2$ test of independence with Yate's Continuity Correction. * $p$ < .01, ** $p$ < .001, two-tailed.

**Performance Trajectories of Classes for the Treatment Group**

Table A.18 shows the growth parameters for the classes (intercepts and slopes). Class 1 contained the highest proportion of participants and had a lower intercept than Class 2. As previously noted, there was a significant difference in intercept values between the classes ($p < .001$). Given that a lower EF performance (*i.e.*, lower error score) indicates superior performance, the Classes 1 and 2, were named the High and Low EF classes, respectively. When considering the slope parameters, Table A.18 shows that both the Low EF and High EF classes showed a significant negative slope ($p = .006$ and $\leq .001$, respectively). Figure A.12 shows the trajectories of the slopes for each class.

The results indicate that the EF performances trajectories increased over the 12-month time interval for the Low EF and High EF classes, with the Low EF class showing the greatest increase (*i.e.*, the greatest reduction in errors) of the two classes. The lowest estimated EF error scores were seen at the 12-month time point. As noted in the method section, these models represent the standard model building process (Jung & Wickrama, 2008). The results were not used to make conclusions about EF improvements following training versus control groups. Such conclusions were drawn from the joint analyses (*i.e.*, of the combined results for the control and treatment groups), described in Chapter 9.

Table A.18.
*Growth Parameter Estimates for the Classes in the Executive Function (EF) 2-Class Model for the Treatment Group*

| **Variable** | **Estimate** | ***SE*** | ***p.*** |
|---|---|---|---|
| 2: Low EF (*n* = 40) | | | |
| Intercept | 104.078 | 4.180 | <.001** |
| Slope | –2.049 | 0.742 | .006* |
| 1: High EF (*n* = 195) | | | |
| Intercept | 49.592 | 1.107 | <.001** |
| Slope | –0.701 | 0.114 | <.001** |

*Note:* 1, 2 indicates model class assignment in model as per Table A.16 and Table A.17. * *p* < .01, ** *p* < .001, two-tailed.

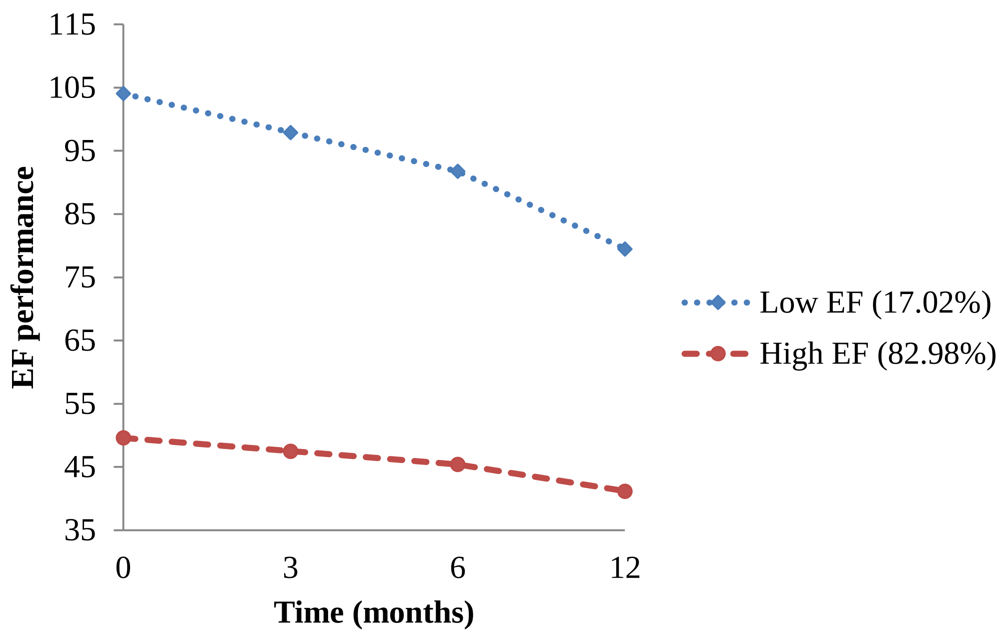


*Figure A.12.* Trajectories of the classes for the executive function (EF) latent class growth analysis (LCGA) across 12 months for the treatment group (*n* = 235).